# SEEING AND THINKING: THE FLEXIBILITY OF VISUAL CONTENT

Nicoletta Orlandi

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the department of Philosophy.

Chapel Hill
2007

Approved by:

William G. Lycan

Jesse J. Prinz

Dorit Bar-On

Ram Neta

Jay F. Rosenberg

ABSTRACT

Nicoletta Orlandi - Seeing and Thinking: the Flexibility of Visual Content
(Under the direction of William G. Lycan and Jesse J. Prinz)

We have a set of seemingly contrasting intuitions about what we see: one is the intuition that the world visually appears to us as a rich panorama of meaningful objects and properties. The second is the idea that what we see is determined by the visual system that we have; this implies both that some things cannot be seen, and that some things can be seen even if we have no conception of them. The third intuition is the idea that the way the world appears to us is also partly determined by our conception of it, and can change given a change in such conception.

I offer an account of the content of visual perception, that is, an account of what is conveyed by our seeing that respects these intuitions and resolves their apparent contrast: I show that a theory that does so is also a theory that is best responsive to the evidence in vision science.

I begin by considering two existing views of visual content held respectively by Jerry Fodor and Paul Churchland. I argue that neither view accommodates the intuitions and, correlatively, is particularly sensitive to the evidence. While differing significantly in their conclusions, both views share substantial assumptions. In line with most of cognitive psychology, they presume that vision is an inferential process, and they further assume that vision has the primary function of grounding our beliefs. I deny both of these shared assumptions. By providing an alternative interpretation of psychological models of vision I

show that we don’t need to think of visual processes as inferential. I further argue that it is best to think of vision as having multiple purposes, producing representational states that can fail to play a justificatory function. Accordingly, I draw a distinction, along the lines of Dretske (1969) between epistemic and non-epistemic seeing.

Together, the rejection of the idea that visual processes are inferential and the distinction between epistemic and non-epistemic seeing make room for a view that resolves the apparent contrast in our pre-theoretical intuitions and that is best responsive to the evidence.

ACKNOWLEDGEMENTS

I thank my committee members for being excellent mentors and just extraordinary people. My advisors Bill Lycan and Jesse Prinz with their patience, wisdom and competence guided me throughout this project. Dorit Bar-On deserves special mention for advising and supporting me at all stages. Ram Neta with his enthusiasm and constructive criticism inspired many sections of this work. Jay Rosenberg, from the very beginning of my graduate career, taught me the most valuable lesson: to read the words on the page.

The following individuals took part in a reading group organized at UNC-Chapel Hill in the Spring of 2006. Most of the ideas in this project were first formulated and discussed there. I thank the participants for their help. In particular: Heather Gert, Bryce Huebner, Doug Long, Dean Pettit, and Dylan Sabo.

Finally, I thank my family – my parents Fiorella and Donato, my brother Gabriele, his wife Valeria and my niece Frida for their love and support.

TABLE OF CONTENTS

# I. INTRODUCTION

We ordinarily think that our visual world is fairly rich comprising meaningful objects and many of their properties. We say we see, for instance, the facing surface of an object as well as its shape and color. We say we see what kind of object it is, e.g. that it is a car, or a tree, or a chair. But we also say we see things that we don't ordinarily think are visible: for example, we say we see a mathematical proof upon looking at one on the board. Or we say we see trouble when looking at a group of drunken adolescents. On second thought, we may admit that mathematical proofs, and troubles are not visible objects: we can conceive or think about them given what we see (what's on the board or the drunken adolescents) but we can't see them. Although there are many things that we can see, there are some things that we can only conceive of or think about given what we see: troubles and mathematical proofs are like that. They are not visible objects, that is, objects that can be detected by our visual system, and what we see certainly depends on what kind of visual system we have.

What we see depends on what kind of visual system we have not only in the sense that the system limits the range of things that we can perceive, but also in the sense that it tells us how the world is independently of our conception of it. There is a sense in which we can see a tree even if we fail to notice that there is a tree, and even if we don't know at all what a tree is. There is a sense in which we and children, who have no concept of what trees are, share the same visual world.

But, while we certainly think that there is independence between what we see and what we conceive, we also tend to think that there is an important relation between them. Our conception of the world can change how we see it: if we have no concept of what a tree is, for instance, then there is a sense in which we don't see a tree upon encounter. Knowing nothing about trees, it does not look to one as though a tree is present. So, acquiring some conception of what a tree is seems to enrich our visual world: it can make it so that it looks to us as though a tree is present when before it only looked to us as though some confusing green and brown "stuff" was present. So, the way the world appears partly depends on the concepts we have.

We then have a set of intuitions about seeing that seem to be in tension with one another: one is the intuition that the world appears to us as a rich panorama of meaningful objects and properties. The second is the idea that what we see is determined by the visual system that we have and this implies, first, that some things cannot be seen, and, secondly, that some things can be seen even if we don't know what they are The third intuition is the idea that although it is true that how the world appears to us is partly determined by the visual system we have, what we see is also partly determined by our conception of the world, and can change given a change in such conception. Whether or not we have some concepts makes a difference to how we see the world. In what follows, I will label this set of intuitions "common sense".

Here, I aim to defend common sense. I aim to give an account of the content of visual perception and of the sense in which such content can change given a change in an observer's conceptual repertoire that preserves the pre-theoretical intuitions outlined above. By the "content of visual perception" I mean what is conveyed to a subject by her seeing. I will assume that, generally, when a subject sees something the world appears a certain way to her: the content of what is seen is typically given by how the world appears. If, for instance, it looks to you as though there is a tree in front of you, then the proposition "there is a tree in front of you" expresses the content of your state. One of the questions I will be addressing is whether our visual system is capable of presenting the world to us in terms of trees or whether it can only give us information about, say, their shape and color from which we infer the presence of trees.

Assuming that visual states are content-bearing states amounts to assuming that vision is representational in the sense that seeing involves being in a state that represents the environment in a certain way, and so in a state that has content. Visual states are states with conditions of semantic evaluation: they can represent correctly or incorrectly. This assumption is hardly questioned in contemporary cognitive science.[1] Vision is usually taken to be a system that, from a set of proprietary stimuli, produces representations of the distal causes of the stimuli that are made available to cognitive systems, like memory, for a number of different tasks (grounding beliefs and judgments, guiding locomotion etc.)[2] How rich or detailed the representations are is a matter of controversy, but that such representations are produced is not. I tend to think that the plausibility of this assumption is a function of how explanatory powerful it is (and has been) in giving models of vision and in explaining the behavior of complex cognitive systems (like us). So long as thinking of vision as representational helps us explain, for instance, why one can be misled by one's visual

[1]J.J. Gibson (Gibson 1979) is famous for questioning it but contemporary Gibsonians, like Alva Noë (Noë 2004) accept that vision is representational.

[2]I will refer to the outputs of vision as "visual representations" and sometimes as "percepts". I will also call a visual perceptual state of which we are aware a "visual experience".

perceptions or how visual states can serve as grounds for our beliefs and judgments, we have little reason to doubt the assumption.[3]

It is important to keep in mind that assuming that vision is representational does not amount to assuming that our access to the world is mediated by the representations. That our visual system produces a representation of the environment that is made available to other cognitive faculties does not imply that such representations are objects to us in the sense that we perceive the world only by perceiving the representations. Representations are theoretical posits used to explain a system's behavior: it is because we are in a state that represents the world in such and such a way that we exhibit such and such behavior. So, if one of the worries for rejecting the idea that vision is representational is a worry about our direct access to the world, the worry can be dispelled.

One of our commonsense intuitions is that the acquisition of conceptual resources brings about a transformation of perceptual content. The way the world appears can change given a change in the observer's conceptual repertoire. This requires saying something about concepts. Here, I will treat concepts as mental representations that subjects possess at times.[4] Further, the possession of a concept typically involves the possession of certain abilities: even if we think of concepts as mental particulars, the ascription of a concept to a subject generally depends on the exhibition of certain abilities on the part of the subject.[5] Now, what

[3]Thus, I will not be concerned with views such as traditional sense-data theory (Russell 1912 and Price 1932). Such views suppose that one only sees, in the sense of being acquainted with, sense-data (in the case of vision, patches of color) and constructs objects out of them. But sense-data are usually not taken to be representational elements: they are mental items that cannot misrepresent.

[4]There is considerable debate concerning what kind of entities concepts are, e.g. whether they are psychological (Fodor 1998) or abstract entities (Peacocke 1992). In line with a good portion of cognitive science, I assume that concepts are mental representations that are ascribed to subjects in order to predict and explain their behavior. But as far as I can see, the arguments that follow do not rest on this assumption and are effective even if we think of concepts as *abstracta*.

kind of abilities? This really depends on how robust an account of concepts one has. At the minimal end of the spectrum, some think that concept ascription is justified if a subject can merely discriminate and recognize something upon encountering it: if you can discriminate a tree and recognize it on multiple occasions, then you can be credited with the possession of the concept "tree". At the opposite end of the spectrum are those who believe that concept ascription is justified only if the subject possesses a language, and is capable of employing a concept in thinking about the world and justifying her beliefs (Sellars 1956; Davidson 1975; Brandom 1994; McDowell 1996).

For present purposes, however, we don't need to decide on this issue. All we need is a preliminary way of understanding the idea that a change in our conceptual repertoire can bring about a change in visual content. We can think of concepts as mental representations that provide a few discriminatory and inferential abilities. The idea is that as one learns new concepts one also learns to see the world differently: as one learns to discriminate, recognize, remember and have thoughts about trees a scene containing trees looks different than it did previously. In what follows, I will speak interchangeably of "concepts", "conceptual repertoire", "knowledge" (or "background knowledge"), "expertise", and "experience". I will also somewhat equate one's "totality of concepts" with one's "theory of the world" (or "background theory"), with one's "conception of the world" and with one's "beliefs" or "system of beliefs". I am aware that there are important differences between these notions but I will follow my interlocutors in using them interchangeably. When necessary I will draw the appropriate distinctions.

[5]Jerry Fodor would deny this requirement on concept possession: possessing a concept, according to him, is possessing a representation, and a representation only involves standing in a nomic relation to that which it represents. I think that there are good reasons to reject Fodor's view but I will not get into the specific arguments. As far as I can tell, the line of argument that I will be pursuing does not beg the question against Fodor on this point.

Attempts to defend pre-theoretical intuitions are sometimes met with scorn in cognitive science. Surely, what is conveyed to us by our visual system and whether it changes with a conceptual change is an empirical issue. How the issue turns out should be quite independent of our intuitions concerning what we see. But, in this case, there are good reasons for wanting to preserve the intuitions. It seems to me that for each of the intuitions there is a suggestive chunk of evidence in cognitive science that a) should be predicted by a theory of visual content, b) is not predicted by the current accounts on offer (at least by the ones I will discuss) and c) it would be predicted if we had a theory that respected the pre-theoretical intuitions. So, wanting such a theory is not only a matter of wanting to preserve common sense, but also a matter of wanting a theory that predicts the evidence.

Take, first, the idea that we visually perceive a relatively rich environment composed of objects and their properties, rather than just, say, patches of color from which we infer the presence of objects. A significant body of evidence shows that children learn to use names of concrete objects, specifically objects belonging to the so-called "basic categories" (see chapter three) before they learn to use words for verbs or for sensory qualities (Rosch 1978; Tomasello et al. 1993). Children are competent in the use of color words, for instance, relatively late in development (around age 4). This seems to suggest that children do not have a developed mastery of what colors are until after they already have a relatively well-developed understanding of what concrete objects are. And since it is plausible to think that our understanding and knowledge of the world partly derives from what our perceptual system makes available to us, it is reasonable to demand that a theory of visual content predicts this fact.

Consider next the intuition that what we see is independent of our concepts and depends on the power and limitations of our visual system. This intuition seems to be the backbone of another intuition we have: the idea that, at least in some cases, we learn some of our concepts from our perceptual encounters so the perceptual encounters should not presuppose the possession of concepts. We learn that the world is a certain way by seeing it a certain way independently of what we antecedently think about it. As it turns out, this is not just an intuition, for there is such a thing as cognitive development. Children and adults differ in what they know about the world and there is both an order to how they learn, and cross-cultural similarities in such order. So, it is reasonable to expect a theory of visual content that predicts this.

Finally, consider the idea that the way the world appears to us is also partly determined by our conception of the world, and can change given a change in such conception. It is fairly well-documented that although children tend to learn names of concrete objects belonging to basic categories (tree, chair, car etc.) and although we tend to more readily identify and describe objects in those terms, a change in expertise can cause a change in how readily we identify and describe the environment. Expert bird watchers, for example tend to display finer discriminatory abilities than ordinary people when it comes to identifying birds, and they tend to use more specific (and/or) more general names in classifying them (Johnson et al. 1997). Cross-cultural studies confirm this evidence by showing that people belonging to different cultures where different contingencies prompt the need to develop different kinds of expertise have a slightly different learning progression (Tardif 1996) and tend to have discriminatory abilities that differ from ours (Boster 1986; Atran 1994). This suggests that different experiences can influence the way the world appears: when we become experts in a certain domain we are able to see aspects of the world that we were previously blind to. And, again, it seems reasonable to expect that a theory of visual content will be able to predict this.

It is my contention that the available accounts of visual content do not accommodate common sense and, correlatively, are not particularly sensitive to the evidence. To simplify the exposition, I consider two broad views of visual content that share substantial assumptions but that reach radically different conclusions concerning what we see. This simplification may seem artificial and there may be substantial differences between views that I am bundling together, but I think that the distinction proves useful in identifying the common presuppositions of each view.

The first broad view of visual content is what I will call “minimalism”. According to minimalism there is a set of things that vision represents, a “visual base”, that is a function merely of having the visual system that we have and that is not affected by our expertise. The way we see the world is fixed while the way we conceive of it can change. Minimalism tends to disallow perceptual development while allowing significant conceptual development. Thus, minimalism is well-suited to accommodate the idea that there is a clear distinction between what we see and what we conceive but less well-suited to explain how our knowledge of the world makes a difference to how we see it. Moreover, depending on the kind of minimalism, this view may not accommodate the intuition that our visual world is relatively rich.

The other view of visual content is what I will call “maximalism”. Maximalism denies that there is a set of things that we see as a function only of contact between our visual system and the world. Visual content is given by the constitutive relations it bears to the concepts we have. Thus, the way the world appears to us depends constitutively on the concepts we have, and can vary with them. There is no visual base, that is, no set of things that we see independently of our theoretical commitments. Maximalism tends to allow a rich view of visual content and to stress the significance of how changes in our background knowledge can affect our perceptual abilities, e.g. our discriminatory abilities. Thus maximalism is well-suited to accommodate the idea that our visual world is rich and the intuition that what we see can change given a change in conceptual resources, but it threatens the independence of what we see from what we believe.

As I have described them, the crucial distinction between the two views is whether they allow a visual base or not, that is, a set of objects that is represented in vision and that is free from theoretical commitment. But although disagreeing on this point, both views agree on the way they understand visual processing. In line with most of cognitive psychology, both views presume that vision is an inferential process where visual representations are produced as a result of an interpretation (Helmholtz 1924-25; Rock 1975; Rock 1983; Gregory1970; Ullman 1979). The stimulus for vision is said to be ambiguous in the sense that a given pattern of stimulation could have been caused by a number of different distal causes and so could give rise to a number of different representations. The visual system is said to solve the problem posed by this ambiguity by performing non-demonstrative inferences where the system bets on the most plausible causes of the stimulus it receives. As we will see, there are good reasons to accept this view of visual processing. For now, it is sufficient to notice that both minimalism and maximalism subscribe to this view. Additionally, they share a tendency to suppose that vision has a primary function or purpose.

Some brands of minimalism, for instance, tend to suppose that the primary function of vision is to enable recognition.

In what follows, I deny both of these shared assumptions. By appeal to the notion of a "natural constraint" (Pylyshyn 1999) I hope to show that vision can solve the problem posed by the ambiguity of sensory states without having to think of it as performing inferences. Natural constraints are usually thought of as principles concerning the physical world that reduce the interpretations that the visual system is allowed to make given the initial data (Ullman 1979, Rock 1983). But although constraints can be thought of as principles, they differ from principles of inference in exhibiting the following features: the constraints are built into the system, their success in producing veridical representations is a function of the environment we evolved in, they are not sensitive to beliefs and judgments, and they don't seem to be available to cognitive systems (Pylyshyn 1999). Thus, by appeal to them we can explain how we get certain kinds of visual representations from ambiguous stimuli without making reference to inferential processes. By denying that vision is inferential, I hope to preserve both the intuition that vision has a relatively rich content and the intuition that how the world appears to us is (at least in one sense) a function of the kind of visual system we have.

Next, I shall deny that vision has a primary function and argue that it is best to think of it as a multi-purpose tool, a tool that enables recognition as well as grounding our beliefs and judgments as well as facilitating locomotion. Thinking of vision this way suggests a distinction between visual states that are available to the conscious subject and visual states that serve other purposes but are not so available. Accordingly, I will draw a distinction, along the lines of Dretske (1969), between two kinds of seeing, and correlatively, two kinds of visual content: epistemic and non-epistemic. The difference between these two kinds of visual content is not a difference in what they represent or in whether they represent at all but a difference in how available and understandable they are to the subject of experience. Seeing epistemically involves being in a visual state that makes a difference to a subject's conscious epistemic life. The subject, for example, can use the visual state to form, and justify beliefs and judgments. Conversely, seeing non-epistemically involves being in a visual state that represents the environment in a certain way but that might not be available to the subject of experience or that is available but still not usable for justificatory purposes. By drawing this distinction, I hope to preserve the intuition that acquiring expertise makes a difference to how we see the world while also holding on to the idea that how we see the world is a function of the kind of visual system we have.

Together, the rejection of the idea that visual processes are inferential and the distinction between epistemic and non-epistemic seeing make room for a view that respects common sense: vision has a relatively rich content representing objects that have characteristic looks in the sense of having characteristic shapes, colors, orientations, and position. In virtue of having characteristic looks a vast variety of objects count as visually represented (trees, pine trees, cars, chairs etc.) Moreover, the view accommodates the intuition that our conceptual repertoire affects how the world appears to us: expertise in a certain domain can enable us to appreciate objects that we already visually represented but that were not available to us for epistemic purposes.

The next three chapters, on minimalism and maximalism, are predominantly negative. I aim to uncover the common presuppositions of the two views and to show that the arguments for them are unsatisfactory. In chapters two and three, I consider two versions of minimalism. Both versions suppose that vision represents a restricted set of objects, a visual base, but one allows for a richer content than the other. The first supposes that the outputs of visual processing are representations of three-dimensional geometric shapes, while the second maintains that the representations are of objects belonging to basic categories (cars, trees, chairs). Both share the assumptions that visual systems perform inferences, and that they are modular. As we will see, the claim of modularity is what allows minimalist views to keep the objects represented in vision at a respectable minimum. In both chapters, I show that the arguments provided are inconclusive. In particular, I present reasons to doubt that perceptual systems are modular, and I show that thinking of vision as having a richer content is better supported by the evidence.

In chapter four, I discuss maximalism and some of its unpalatable consequences. Maximalism shares with minimalism the assumption that vision is inferential but it rejects modularity. I hope to show that, while the rejection of modularity is justified, the arguments for maximalism are also unsuccessful: defenders of maximalism tend to overlook the notion of constraint that vision scientists often use in explaining how visual processes occur and overemphasize evidence on cross-cultural differences in perceptual reports and discriminatory skills. In concluding this chapter, I suggest that we should draw a distinction between epistemic and non-epistemic seeing, that is, a distinction between states that are available to the conscious subject for epistemic purposes and states that are not.

In chapters five and six, I develop my positive account of visual content. In chapter five, I show that it is plausible to think of visual systems not as inferential but as governed by constraints that have been built into the systems. I also provide an account of what is involved in aspect shifts that paradigmatically occur when we see multi-stable figures like the famous duck-rabbit. Such cases are often taken to be more naturally explained by inferential accounts of vision: the same figure is seen in multiple ways, and this is taken to be due to a change in interpretation. I show that cases of vision in the presence of ambiguous figures can be explained in a way that does not make reference to inferences. According to my view, aspect-shifts are eventuated by shifts in attention that allow one to see the same object in new ways. Seeing the duck-rabbit as the figure of a duck is not a matter of deploying the concept "duck" to interpret the figure but a matter of paying attention to the visible features of the figure that prompt the perception of a duck-figure. We see the figure as a duck-figure because the figure is duck-shaped and not because we think of it that way.

In chapter six, I go back to the distinction between epistemic and non-epistemic seeing and explain how this distinction helps in preserving the intuition that our concepts partly determine how we see the world. I suggest that perceptual development is the process of moving from non-epistemic to epistemic perceptual states and offer a model of how this process may take place. Borrowing from research in developmental psychology (Karmiloff-Smith 1992) I suggest that we can think of the process of perceptual development as a process where representations become more and more available to the conscious subject and integrated in a network of others. Unlike Dretske, I don't think of the move from non-epistemic to epistemic seeing as a simple move from analog to digital form, and I don't think of the move as a move from a perceptual to a cognitive state. I aim at supplementing the approach in Dretske by drawing a distinction, lost on his account, between concepts and high-level perceptual representations that are not yet conceptual. I argue that epistemically seeing something is being in a state that is available to the subject for epistemic purposes and that may be influenced by one's expertise but that is still essentially a perceptual state. I conclude by showing how my view, as opposed to both minimalism and maximalism, preserves common sense.

# II. MINIMALISM1: 3D SHAPES

In this chapter, I argue against one type of what I called minimalist accounts of visual content. Minimalist accounts of visual content hold that vision represents a set of objects as a function merely of contact between our visual system and the environment. The particular kind of minimalism that I discuss in this chapter (Minimalism1) holds that vision has a relatively impoverished content. Contrary to what we ordinarily think we see, we don't see much: for instance, we don't see full-fledged objects like trees, cars and chairs. We rather see their shape, and position from which we infer or construct or think of the presence of trees, cars, and chairs. The impression that our visual world is rich is an appearance: most of what we think we see we really just conceive given what we see. Thus, this version of minimalism puts into questions two tenents of common sense: one is the idea that our visual world is relatively rich, and the other is the idea that the way we see the world changes with a change in expertise.

Different reasons have traditionally been given for thinking that minimalism1 is true. In this chapter, I survey some of the most popular reasons derived from cognitive science and argue that they are not satisfactory. I believe that vision has a richer content than what minimalism supposes, but instead of arguing directly for my conclusion the strategy is to argue that the reasons for thinking that vision has a relatively impoverished content are unsuccessful. This way of proceeding may appear immediately unsatisfactory: the fact that the arguments that I consider in favor of minimalism1 are no good does not imply that there couldn't be other good arguments or that minimalism1 is false. But I hope that the present chapter, when integrated with the evidence and analysis provided in later chapters will constitute a much more convincing refutation of minimalism1. So, chapter two should be read as a first-step toward a conclusion that will be reached only later.

Here, I analyze in particular David Marr's theory of visual perception and his reasons for thinking that vision represents only volumetric shapes rather than full-fledged objects. I do so for two reasons: one is that Marr's work was and is particularly influential in touting the ideas that vision is inferential (although it is matter of controversy in what sense Marr thought that vision in inferential) and that it is modular. So, analyzing his view offers an excuse to get clearer on what modularity amounts to, and on how visual inferences are (supposedly) carried out. Since modularity can be independently used to argue for a relatively impoverished view of visual content, I show, following Prinz (2006a), that the evidence in vision science does not support modularity and that it in fact suggests that visual systems are non modular.

The second reason why I think it is important to analyze Marr's view is that his main motivation for thinking that vision represents in particular three-dimensional shapes is also fairly popular and has inspired others (Pylyshyn 1999) to hold the same view. Marr appeals to cases of associative agnosia to argue that the computation of shape is quite independent from the recognition of objects. By showing that cases of associative agnosia do not show what Marr thinks they show I hope to deprive minimalism1 of a good source of evidence.

In the opening section of this chapter, I present the idea that visual processes are inferential, and explain the reasons behind the idea. I then move on to discuss modularity and why modularity can be seen to provide reasons for an impoverished view of visual content. I then consider Marr's specific reasons for thinking that visual systems output representations of geometric shapes and argue that they are not satisfactory. I conclude by showing that the evidence in favor of modularity is inconclusive and that we have reasons to suppose that visual systems are non-modular.[6]

**The Inferential Nature of Vision**

Vision, according to Marr, is the process of discovering, from images, what is present in the world, and where it is. Supporting this simple characterization is a widely accepted view of how vision works (Rock 1975, Rock 1983, Gregory1970). In line with most of cognitive psychology, Marr thinks of Visual processes as operations over representations where more abstract representations are built out of less abstract ones. The initial representations are provided by mechanisms of transduction that are responsible for producing representations of the proximal stimulus: in the case of vision the proximal stimulus is constituted by wavelengths of light. Transducers are proprietary to a given sense modality: in vision, the mechanisms of transduction are retinal photoreceptors that produce representations of light intensity values. The initial representations produced by transducers are also the initial data for the visual system and contain information about contrast, acuity,

[6]Another type of minimalism that cannot be appropriately mapped into the divisions made in the introduction is gaining some popularity. I think of it as a kind of minimalism because it maintains that, contrary to what we ordinarily think, we do not visually perceive a rich and detailed environment by simply staring at it (Noë 2004). Seeing the environment essentially involves being situated in it and finding out about it. Recent experiments on change blindness have been taken to support this view by showing that the visual system does not construct detailed representations of the world: the representations are rather "gappy" and piecemeal, and accordingly, our access to the world, if not integrated by action, is only gappy and piecemeal. In order to preserve argumentative clarity, I do not discuss this position here. Let me just note that, as far as I can tell, the claims that characterize this view of visual content are orthogonal to my argument and to the views that I do discuss. It may well be true that we need to do more than just stare in order to perceive a detailed environment but the question still remains: once we have done more, in what terms do we see it?

line orientation and color. These are sometimes referred to as “sensory data” or “sensory states”. The visual system is responsible for converting the initial representations of proximal stimuli into representations of the distal causes of the stimuli: such representations are made available to other cognitive faculties for a number of different tasks (moving, forming beliefs, storing in memory etc.) So, vision moves from retinal images to detailed representations of the distal causes of the retinal images.

Now, it is widely accepted that ambiguity is the hallmark of sensory states (Frisby 1980; Rock 1975; Purves et al. 1997). A two-dimensional pattern of light on the retina underdetermines its distal causes, in the sense that the pattern could have been caused by a number of three-dimensional things in the environment. Yet in almost all cases we attain a unique percept (reversible figures are a notable exception and we will talk about them in chapter 5). Visual processes are then confronted with the problem of interpreting sensory data to produce representations of the distal cause. Richard Gregory, for instance, describes the problem of vision as follows (Gregory 1970, p.25):

> “At this point we might be tempted into thinking that perception is simply a matter of combining activity from various pattern-detecting systems, to build up neural descriptions of surrounding objects. But perception cannot be anything like as simple, if only because of a basic problem confronting the perceptual brain – the ambiguity of sensory data. The same data can always ‘mean’ any of several alternative objects. But we experience but one and generally correctly. Clearly there is more to it than the putting together of neurally represented patterns to build perceptions, for decisions are required. We should look at the ambiguity of objects to see this more clearly. Establishing that a given region of pattern represents an object and not background is only a first step in the perceptual process. We are left with the fatal decision: What (kind of) object is this?
> The problem is actual because any two-dimensional image could represent an infinity of possible three-dimensional shapes. Often there are extra sources of information available; for example stereoscopic vision, or changing parallax as the head moves, but the fact remains that we can nearly always arrive at a reasonably reliable solution to the problem: “What object is this?” Even though, the number of possibilities is infinite.” (p.25)

In order to solve the problem posed by sensory states visual systems are often said to perform non-demonstrative inferences where the visual system bets on the most plausible causes of the sensory stimulus it receives. The inferences are usually said to be performed subconsciously (Von Helmholtz 1924-25) or sub personally (Rock 1983) and to produce detailed representations of the environment. Since the inferences tend to produce a unique percept, it is common to think that they are executed by using knowledge of the world that the visual system already has (Bruner 1957). Such knowledge is supposed to constraint the inferences that the visual system is allowed to make. In particular, the inferences are brought out by conforming to principles concerning the physical make-up of objects in the environment (Marr 1982; Ullman 1979; Spelke 1988; Rock 1983). The visual system organizes the initial information into representations of objects by assuming, for example, that the causes of visual stimulation are rigid and so that the interpretation "rigid structure" should be preferred to a competing interpretation. Given some initial data, the visual system tends to produce representations of rigid structures rather than of non-rigid ones because the system knows that objects in the environment are typically rigid. In this picture, vision is an inferential process that starts from some initial representations (the premises) and ends with more abstract representations (the conclusions) by conforming to some principles of inference.

It is reasonable to wonder about this characterization of visual processes but acceptance of the idea that they are inferential is widespread and the lack of alternatives is one motivation for such acceptance. In chapter five, I will argue that there is a plausible alternative. For now, let me point out that there is a worry prompted by the idea that visual systems are inferential. The worry is about the compatibility of accepting this view while also holding on to minimalism. As I have described it, minimalism about visual content holds that vision represents a restricted set of objects and that it represents them as a function merely of contact between our visual system and the environment. But if we understand visual systems as inferential in the way outlined above, then the representations produced by vision are not a function of a “mere contact” between our visual system and the environment. They are rather a function of inferential processes that use knowledge of the world to produce representations of it. Thus, the representations, while being the product of visual processing alone, are not free from theoretical commitment. It is then reasonable to suppose that it is in the minimalist’s best interest to find an alternative explanation of how visual processes take place.

Some minimalists, however, think that this alternative explanation is not necessary (Fodor 1988). All that’s required is a commitment to the idea that visual systems are modular: they are innately specified systems that have a very restricted knowledge of the world and that work sub personally and independently of the subject’s control. Thus, although there may be a sense in which their outputs are theory-laden, there is also a sense in which they are not: they employ innate knowledge that is not available to the conscious subject and they are not influenced by the conscious subject’s theory of the world. It seems then, that acceptance of modularity is an important component of being a minimalist. As I will explain next, modularity licenses the minimalist to preserve a relatively impoverished and theory-neutral view of visual content.

**Modularity**

Marr embraces the idea that visual systems, *qua* "input systems", are modular (Marr 1982, p. 102). To be modular is to exhibit a number of features that allow the systems to act on a given input (Fodor 1983). A few of these features are particularly relevant to the issue of visual content. Two of them are localization and characteristic breakdowns: modules are localized in dedicated areas of the brain and they can be selectively impaired. This means that focal brain lesions cause selective deficits: damage to some parts of the brain impairs some mental abilities without impairing others. Visual agnosia, to which we will return, is taken to be one of these deficits.

Further, modules are informationally encapsulated, inaccessible, fast, and their outputs are mandatory and shallow.[7] Informational encapsulation amounts to the idea that visual mechanisms have access to only a limited amount of information in performing their operations, e.g. they do not employ information stored in memory. So, if we take vision to be modular, then the visual system takes the representations produced by retinal transducers and transforms them into representations of distal causes without employing the subject's general knowledge of the world. Modules have access to only a proprietary set of perceptual concepts in performing their inferential operations (Pylyshyn 1999). They may have access, for example, to concepts like "three—dimensionality" and "rigidity", without having access to richer information about the use and purpose of objects in the environment. Since modules employ only a proprietary vocabulary in performing their operations, their outputs will likely be simple.

[7] The remaining features of modules are their ontogenetical determination – the fact that they develop with a characteristic pace and sequence – and their domain specificity – the fact that they have a set of proprietary inputs. These two features are pivotal to the claim that modules are innately specified. I will not critically evaluate these two features here, because they do not appear to be central to the issue of visual content: Prinz offers a convincing criticism in Prinz (2006a).

Inaccessibility is the other side of encapsulation: while encapsulation essentially means that not much information is let in, inaccessibility implies that not much information is let out. Modular systems make contact with cognitive systems at only one level. We don't have introspective access, for example, to the representations produced by transducers.

Additionally, visual processes are fast and their outputs are mandatory: modules generate their outputs quickly and automatically. Little processing is required to move from the initial sensory representations to representations of the distal causes, and what these representations represent escapes the observer's control. One can fail to judge that the world appears as vision presents it to us, but one can hardly change the way the world appears. We can't help seeing a visual array as consisting of objects distributed in three-dimensional space, even though we can help judging that the space actually contains the three-dimensional objects we perceive (Fodor 1988).

Thinking of visual systems as modular favors the idea that the outputs of visual processing are 'shallow' in the sense of being relatively simple. Since visual systems do not have access to a lot of information in performing their operations and they process sensory data quickly and automatically, what they produce must be the product of little processing and so it must be relatively simple. Additionally, inaccessibility suggests that perceptual systems make contact with cognitive systems at only one level. Together, these features of the modules constitute a reason for minimalism: visual representations cannot be too rich and they are representation of items at a single level of abstraction, the level of modular output. But notice that what exactly this means remains an open question: that is, it is still an open question what level of abstraction is the level of modular outputs and exactly how simple the representations at that level have to be. In the next section, I consider Marr's answer to these questions and argue that his answer is unsatisfactory.

**Three-Dimensional Shapes and Associative Agnosia**

Supposing that visual processes are inferential and modular leaves open the question of what kind of representations they output. Marr's answer to this question is that visual input systems output representations of object-centered, three-dimensional shapes that are position invariant and that encode the salient geometric features and orientation of objects. According to Marr, the visual system proceeds in, roughly, hierarchical order: first, contrast, acuity, orientation and color are detected and somewhat organized into a two dimensional representation (this is the 'primal sketch'); then a viewer-centered representation is achieved by incorporating information about color, surface and depth (the 2.5 D sketch), and finally a volumetric and geometrical representation of the shape of an object is generated (the 3 D sketch). [8] If this is true then vision has a relatively impoverished content: it represents the volumetric shape of, say, a tree, but not a tree itself.

Marr's generic reason for thinking that vision outputs representations of volumetric shapes is that the outputs of visual processing should be descriptions of the environment that are useful to the organism. Volumetric shapes that are position invariant, Marr thinks, are useful in making recognition of objects possible when matched to an observer's general knowledge of the world. The idea is that object recognition is achieved in roughly two stages:

[8] Marr is not always clear on what he takes the output of visual processing to be. At the beginning of chapter 4 he says: "The construction of the 2 ½ D sketch is a pivotal point for the theory, marking the last step before a surface's interpretation and the end, perhaps, of pure perception." At the very least, Marr thinks that the 3 D sketch is not the only stage of visual processing that is introspectively accessible to the subject and this raises questions for Marr's commitment to modularity.

we first see the shape and orientation of an object and then associate our general knowledge of the object with the visual percept. As in a Von Neumann computer, visual recognition requires that memory be searched for a representation that resembles the current perceptual representation. The current and stored representations are compared using an explicit comparison process that is itself part of the program of the computer. When a match is found, the knowledge of the object associated with the representation stored in memory is made available to the system. In this picture, we first see, say, the shape and orientation of a tree and then match such representation with a stored and "semantically rich" representation of a tree achieving recognition.

That this is Marr's specific reason for thinking that visual processing outputs representations of shapes that need to be matched to some background knowledge is confirmed by his appeal to cases of associative agnosia that, incidentally, are also used as evidence for the view that visual systems are modular (Marr, 1982, p. 35). Associative agnosia is a selective deficit that can be caused by a number of brain lesions (Farah, 2004 p.88). One of its most interesting features is the dissociation of recognition from other more basic visual capacities. Patients affected by associative agnosia are generally able to see the shape of common objects but unable to recognize their name and their "semantics" – that is, their use and purpose, how big they are, how much they weigh, what they are typically made of, and so forth. Particularly impressive is the patients' capacity to draw accurate line-drawings of the objects seen (Farah et al., 1988): such capacity suggests that visual perception is intact, or at least adequate to the task of recognizing objects. Since such capacity is intact, the other sense-modalities are intact, and the subjects' conceptual knowledge of the objects is intact something else is thought to be responsible for the condition. What's thought to be responsible is the inability to match the visual percept with a semantically rich representation of objects stored in memory. In other words, the condition is caused by the inability to match the visual percept to the rest of the subject's conceptual knowledge: thus, recognition fails because the percept is not subsumed under the relevant concept.

Cases of associative agnosia are taken by Marr to show that a) the representation of the shape of an object is a fairly independent process from the representation of its use and purpose and b) vision alone can deliver an internal description of the shape of a viewed object even when the object is not recognized. This, in turn, is taken by Marr to suggest that the primary purpose of vision is to produce object-centered representations of shapes that make recognition possible. Marr's overall acceptance of modularity contributes to making this picture plausible. In the following section, I present reasons to think that visual systems are non-modular. In this section, I want to raise some doubts about whether cases of associative agnosia show what they are often taken to show.

The dominant explanation of associative agnosia has it that cases of agnosia are cases of patients with normal vision but abnormal matching of visual percepts to background knowledge. This is supposedly shown by their capacity to draw accurate outlines of objects that they cannot recognize. But there are a number of considerations that militate against this way of interpreting the cases. The first is the fact that the drawings are achieved with a lot of difficulty and by using a piecemeal strategy. The impression that agnostic patients have intact shape perception but damaged matching can be preserved only if we think of them as smoothly outlining the objects that they can't recognize: but this is far from true. The drawings are achieved in strikingly abnormal ways by carefully copying an object line-by-line and by often having to keep track of the lines in the drawing by tracing them with a finger (Farah 2004). Additionally, the recognitional mistakes of agnostic patients tend to be very sensitive to the quality of the stimulus they receive. Their inability to recognize objects is much more pronounced when they see the objects in a photograph or in an outline, suggesting that their deficits, is highly dependent on the richness of the stimulus rather than on the lack of some higher order cognitive connection. Finally, the kind of mistakes they make when they misidentify an object are telling: patients tend to mistake an object for another object of roughly similar shape but of different dimensions (a key with a tool of some sort, a can opener with a key, a baseball bat with a knife, or a thermometer). If we think that representing the shape of an object also involves representing its rough size, this suggests that shape perception itself is not intact.

If this is true, then cases of associative agnosia cannot be taken alone to show that vision represents only geometric shapes that make recognition possible. The representation of the shape of an object may not be dissociable from the representation of the object itself and its meaning. Now, one may protest that Marr's evidence for his view is not just constituted by cases of associative agnosia, but by his overall acceptance of modularity. The evidence that visual systems are modular seems overwhelming, and modularity can be independently used to argue that modular outputs have to be simple. Geometric shapes, given their simplicity, are a good candidate. In the next section, I argue that the evidence for modularity has been overstated and that visual systems do not exhibit the features that modular systems are generally taken to have. This should constitute an argument not only against Marr's specific position but also against any minimalist position that appeals to modularity to make the case.

**Questioning Modularity**

In the previous section, we saw that Marr can be interpreted to have two reasons for minimalism about visual content: his acceptance of modularity and his appeal to cases of associative agnosia. The two reasons are related: cases of agnosia are often cited as evidence for modularity, and *vice versa* modularity is said to predict cases of agnosia. Since we have seen that cases of agnosia are unconvincing, in this section I consider whether we have good reasons to think that visual systems are modular. More specifically, I consider whether we have good reasons to think that visual systems have the features of modules that are relevant to claiming that their outputs are relatively simple. I argue that we don't. Given the evidence, there are serious doubts concerning the adequacy of modularity as a description of visual systems. I consider most of the central features of modules discussed in this chapter, and show that, for each of them, there is evidence to show that the visual system doesn't have that feature.

If visual systems are modular then they should produce mandatory outputs and they should be localized, have characteristic breakdowns, be informationally encapsulated and inaccessible. Consider first localization and characteristic breakdowns. The evidence that visual systems have these two features may seem overwhelming: the primary visual cortex or V1 situated in the occipital lobe is deemed responsible for most basic visual processing. But there is, in fact, considerable disagreement across studies and laboratories concerning the precise location of areas of the cortex responsible for processing visual data (Uttal 2001). For instance, it is not clear what part of the cortex is responsible for processing color and motion

(Prinz 2006a). Further, visual functions activate broadly distributed regions of the brain. By concentrating on regions of more intense activation researchers have often overlooked other areas that play an important role in the performance of a given function, thereby assuming rather than proving localization. Recognizing target faces, for example, produces responses in widely distributed regions that include the occipital, parietal, temporal and frontal areas (Jiang et al. 2000). And visually perceiving a simple pattern of dots produces responses in three different regions of the cortex – V1, V5 and the posterior parietal region (Büchel et al. 1998). Moreover, the same cortical area is often associated with different functions: the Fusiform Face Area that is commonly thought to be dedicated only to face recognition (Kanwisher 2000) is also likely to be essentially involved in the recognition of other objects, like cars and birds (Gauthier et al. 1999; Gauthier et al., 2000).

Lesion studies are open to similar problems. Too often research on various kinds of visual agnosia is based on individual studies (Ellis & Young 1988). Moreover, the same deficit can be associated with lesions in different parts of the brain: simultagnosia, a condition characterized by the inability of patients to perceive an object or scene in its entirety can be caused by lesions to the parietal lobes, the occipital lobes or the posterior temporal cortex (Farah, 2004). Finally, visual impairments are often accompanied by other kinds of impairment: simultagnosia, for instance, is invariably associated with the inability to read (Farah 2004).

This evidence suggests that, although it may be true that the brain is divided into parts that perform different functions (hardly anyone denies that), the stronger claim that certain regions are exclusively dedicated to the performance of certain tasks is not corroborated. The same area of the visual cortex can perform multiple functions and the same function can be performed by different parts of the brain.

Next, consider the idea that visual systems are encapsulated and that their outputs are mandatory. Visual illusions are often cited as evidence for these two features. One of the characteristics of visual illusions is that our knowledge that they are illusions does not change their appearance: we can't help seeing the two lines in the Müller-Lyer illusion as different in length even when we know that they are not. This is taken to show that our background beliefs about appearances do not change them, and so that we have little voluntary control over how vision presents the world to us. But there are at least two problems with this kind of argument. One is the consideration that alternative explanations can be given of the apparent rigidity of visual illusions. Prinz, for instance, argues that visual illusions only show that when perceptions and beliefs come into conflict perceptions trump beliefs and this is useful given that perceptions are often called to serve in the adjudication of clashes of opinion (Prinz 2006a).

Secondly, some visual illusions are reversible. The Herring illusion, for example, can be reversed by focusing on the vertex in the center of the figure and by trying to deliberately see the vertex three-dimensionally. If the focus stays on the vertex and the parallel lines become peripheral in the field of vision, the lines appear straight (Ihde 1977). So, the idea that visual illusions represent strong evidence for encapsulation given that they cannot be reversed is actually false. Even in the Muller-Lyer, case one can experience a perceptual change when informed of the nature of the illusion. The most popular explanation of the illusion attributes it to size constancy (Rock 1983). Although we are often only aware of perceiving two lines of unequal length, our visual system tends to perceive the lines as two edges: the inside edge of a box further away from the observer and the outside edge of a box closer to the observer. [9] Since the two edges are perceived to be at different distances, the visual system adjusts their size for size constancy (the assumption that objects in the environment do not change their size as they move) and makes one appear bigger than the other. Although it is true that it is very hard to perceive the two lines as equal in length even after we learn that they are it is equally true that upon hearing this explanation it is possible to perceive the illusion as made up of edges at different distances rather than as lines. And it seems that this *is* a change in the appearance of the illusion prompted by a piece of background knowledge.

Additionally, ample empirical evidence suggests that our knowledge does affect our visual perceptions. Linguistic categories, for example, can affect something as basic as color perception: speakers of Russian who employ different names for lighter and darker shades of blue are better in a number of discriminatory tasks than English speakers who do not make such distinction (Winawer et al. 2007). Our previous knowledge of the color of various objects has been shown to affect our perceptual judgments of color (Bruner et al. 1951) and our knowledge of size is known to affect our perceptual judgments of distance (Hastorf 1950). Neurological evidence indicates that practice in discriminating small motions alters significantly brain potentials in the primary visual cortex suggesting that practice affects very early stages of visual processing (Fahle et al.1996).

Finally, consider inaccessibility, the idea that visual systems make contact with cognitive systems at only one level. This idea is said to be supported by the observation that we do not have conscious access to the workings of the visual system. And this seems right:

[9]This explanation is confirmed by studies on visual agnosia. Patients with severe impairment in depth perception are less sensitive to the illusion (Turnbull et al. 2004).

we do not have conscious access to the early representations of retinal stimuli (supposing that there are any) or to how the visual system achieves size constancy. But this only shows inaccessibility to the subject of experience not to cognitive faculties in general. As I argue in the next chapter, it is important to keep in mind that there is a distinction between cognition and conscious thought: it is perfectly plausible to suppose that visual systems make available to cognitive faculties representations that are not available to conscious thought. It is plausible, for instance, to suppose that vision produces representations that get stored in memory without us being aware of such representations. There is evidence to suggest that this is the case (Thornton et al. 2000): observers are often very poor at reporting changes in their visual environment (change blindness) but the changes seem to be represented at a level below awareness. Observers perform above chance in identifying what change occurred in a scene even when they cannot report the change.

If what I have said in this section is right, then we should have serious reservation about thinking of visual systems as modular. Or, better, we should have serious reservations about thinking that the features of modules that support minimalism1 are also features of visual systems. Visual processing doesn't seem to be encapsulated, inaccessible, mandatory, localized and subject to characteristic breakdowns. But then we are back to consider a worry that we expressed in the opening section of this chapter. The worry was about the compatibility of accepting the view that visual processes are inferential while also holding on to minimalism. Minimalists about visual content hold that vision represents a restricted set of objects and that it represents them as a function merely of contact between our visual system and the environment. The particular branch of minimalism I discussed in this chapter adds to these theses, the further claim that visual content is relatively poor. But in order to hold on to these theses and, at the same time, preserve the idea that vision is inferential in the way described, the minimalist1 has to commit to modularity. Since we now have reasons to doubt that visual systems are modular we are left wondering whether minimalism should be given up or whether we should get rid of one of its theses while preserving the others. In chapter three, I discuss and critically evaluate a minimalist account that allows for a richer view of visual content.

**Conclusion**

In this chapter, I argued against minimalist accounts of visual content in general by arguing against the view that visual systems are modular, and against Marr's specific branch of minimalism by questioning the evidence from associative agnosia.

The argument against modularity poses a specific problem for the minimalist. Since minimalists tend to suppose that visual processes are inferential, they have to commit to modularity in order to keep what's represented in vision at a respectable minimum, and in order to hold that what is represented is not determined by the observer's overall theory of the world. But now that modularity has been put into question, we are left wondering about the fate of minimalism.

In the next chapter, I consider a kind of minimalism that allows for a richer notion of visual content.  Specifically, I consider Jerry Fodor's idea that visual systems represent objects belonging to basic categories: cars, trees, chairs etc. Fodor's view has much in common with Marr's. Like Marr, Fodor thinks of visual systems as operations over representations. Like Marr, Fodor thinks of visual processes as inferential and modular, thus also committing to the idea that vision makes contact with cognition at a single level. But unlike Marr, Fodor allows for a richer visual content. Since Fodor provides independent evidence for his view, it is worth considering it. Perhaps minimalism can be saved by simply admitting that vision has a richer content than minimalism1 supposes. As we will see, this is not the case. I argue that Fodor's evidence does not support his view and that if we adopt Fodor's own criterion we end up committing to a very different picture of perceptual outputs, a picture that Fodor would reject. This alternative picture is the main topic of chapter four.

## III. MINIMALISM2: THE REVENGE OF THE GIVEN[10]

In the previous chapter, we looked at an account of visual content that questions common sense in supposing that we see less than what we ordinarily think we see. A rigorous scientific understanding of visual systems shows that vision represents a lot less than what the layman supposes. Most of what we ordinarily say we see (trees, cars, chairs) is the product of our rich stored knowledge of the world rather than of perceptual processing alone. In this chapter, I consider an account that, although subscribing to some tenents of minimalism1 agrees with common sense in supposing that vision has a richer content. The particular kind of minimalism I consider is one that accepts the view that vision represents a restricted set of items and that it represents them in a way that is not affected by the subject's overall knowledge of the world. But unlike minimalism1, minimalism2 supposes that vision represents more than geometric shapes: it represents objects and in particular middle-sized objects belonging to basic categories. Objects belonging to basic categories have to be distinguished from objects belonging to more abstract or more specific categories: a tree, for instance, is an object belonging to a basic category whereas a plant is an object belonging to a more abstract category and a pine tree is an object belonging to a more specific category. Minimalism2 maintains that visual systems output representations at the single level of abstraction of basic categories. Objects belonging to more specific or more general ones are

[10]"The Revenge of the Given" is the title of a recent and yet unpublished paper by Jerry Fodor (2006). In it, Fodor seems to revise the position he held in the "Modularity of Mind" (discussed here). Fodor is now inclined to think that the given, constituted, for him, by unconceptualized representations, may be inaccessible either to voluntary report or to cognitive processes that aren't perceptual, or to both (p.2). I doubt that Fodor provides a convincing argument for such representations, but I have no space to argue for it here.

objects that we can come to *conceptually* represent given the acquisition of expertise, but they are not visually represented. We see cars, trees and chairs but we can only think of sports cars, pine trees and kitchen chairs.

One of the motivations for holding this view of visual content is again the idea that visual systems are modular. Since I provided a criticism of modularity, I don't repeat it here. I rather focus on the additional argument adduced for the position in question. I argue that the argument does not support the conclusion and that if we accept it, we end up with a very different position concerning visual content.

**Basic Objects**

Fodor wants to preserve the intuition that there is a clear distinction between perception and cognition while also holding on to the idea that visual processes are inferential. He does so by committing to modularity. One would then expect his position to be similar to Marr's: the representations outputted by the modules are object-centered representations of geometrical shapes, that when combined with stored representations of objects make recognition possible. But in the *Modularity of Mind*, Fodor suggests a different picture of modular outputs. Although he admits that his remarks on the subject are "highly speculative", he rejects the idea that visual modules output simple representations of geometrical shapes.[11] He says:

> "Moreover, various candidates that satisfy the shallowness test […] must nevertheless be rejected on grounds of phenomenological inaccessibility. I am thinking of such representations as Marr's 'primal', '2.5 D', and '3D' sketch. Such representations are certainly shallow enough. Indeed, they would seem to be too shallow. If we accept them as defining visual processor outputs, we shall have to say that even object

[11] Fodor alludes to this position also in Fodor and Pylyshyn 1981.

> recognition is not, strictly speaking, a phenomenon of visual perception, since, at these levels of representation, only certain geometric properties of the stimulus are specified. But, surely, from the point of view of phenomenological accessibility, perception is above all the recognition of objects and events." (Fodor 1983, pp. 93-94)

And in the footnote to the paragraph just quoted Fodor adds:

> "It may be thought Pickwickian, after all that we've been through together, for me to cleave to phenomenological accessibility as a criterion of the output of the visual processor. I must confess to being influenced, in part, by ulterior – specifically, epistemological – motives. It seems to me that we want a notion of perceptual process that makes the deliverances of perception available as the premises of *conscious* decisions and inferences; for it seems to me indubitable that, e.g. it sometimes happens that I look out the window, see that it is raining, and decide, in light of what I see, to carry my umbrella. If we allow that the deliverances of the input system are *very* shallow representations (edges and colors, say) then we shall have to hold either that input analysis is a very much less rich process than perceiving – mere psychophysics, in effect – or that the intuition that one *sees* such things as that it's raining – and the rain – is misled. Since I feel no inclination towards either of these alternatives, I want a vocabulary for the output of the visual processor which specifies stimulus properties that are phenomenologically accessible and that are, by preference, reasonably close to those stimulus properties that we pretheoretically suppose to be visible." (Fodor 1983, p. 136 footnote 31)

Fodor wants the outputs of the modules to be simple enough to be plausibly produced by encapsulated and automatic processes, but not simple enough to be phenomenologically inaccessible to the subject, where phenomenological accessibility is roughly, accessibility without sustained inspection (Fodor 1983, p.96). The representations outputted by the modules have to be simple, but they also have to represent things that a subject would report if she were to spontaneously describe what she is seeing, and make decisions on that basis. In line with this idea, Fodor argues that visual-input systems output representations of basic categories, e.g. car, tree, chair etc. Basic categories are the categories in terms of which people find it more natural to recognize and identify things, and so, Fodor says, they are likely to be the categories in terms of which people tend to see the world.

To support this view, Fodor cites a number of studies conducted by Eleanor Rosch designed to show that there is a level of categorization that is informationally rich, and so it is the categorization that people tend to use to identify objects. The world, Rosch says, “consists in a virtually infinite number of discriminably different stimuli. One of the most basic functions of organisms is the cutting up of the environment into classifications by which non-identical stimuli can be treated as equivalent.” (Rosch et al. 1976, p382)

In a series of experiments, Rosch argues for a single level of categorization that carries the most information, and possesses the highest cue validity. There are categories that in providing a good deal of information about the objects contained in them help us group together similar objects and distinguish them from dissimilar ones. So, basic categories like “tree”, “chair” and “car” are useful in cutting up the environment because they define categories where objects share the most attributes; more abstract categories like “furniture” or “vehicle” are not as useful as the objects in them share only a few attributes; and more specific categories, like “kitchen chair” and ‘sports car’, are not as useful either because the objects in them share too many attributes with objects in other categories (for example, “kitchen chair” shares most of its attributes with other kinds of chairs). Given this usefulness of basic categories it makes sense to hypothesize that subjects will use such categories rather than more abstract or more specific ones in recognizing and identifying objects: that is, subjects will be more inclined to describe what they see in terms of chairs and cars rather than in terms of furniture and vehicle.

This hypothesis is confirmed by several experiments. In one of the experiments subjects are asked to indicate whether pictures of objects are correctly or falsely designated by names presented immediately prior to the pictures. The results show that subjects tend to designate objects with basic category names more rapidly than with names of more abstract or more specific categories. This indicates that objects are recognized as members of basic categories faster than as more abstract or more specific ones. In other experiments, subjects are asked to spontaneously identify objects: the results, again, confirm that the basic object names are the lexical items normally chosen to refer to things, and so that the basic categories are those in terms of which people spontaneously report what they perceive. Finally, names of objects belonging to basic categories are learnt by children earlier than names of objects belonging to other categories suggesting that basic objects are easily accessible to us.

Fodor takes these experimental results to suggest that modules make available to cognitive faculties a more abstract level of representation than what Marr supposes. He says:

> "Basic categories are phenomenologically *given*; they provide, as it were, the natural level for describing things *to oneself*. A glance out the window thus reveals; a lady walking a dog, rather than a lady walking a silver-grey, miniature…etc. (Of course, sustained inspection alters all this. But phenomenological salience is accessibility *without* sustained inspection)." (Fodor, 1983, p.96)

And in turn he takes this basic level to be the level of output of the modules.

The line of reasoning that Fodor is using seems to be the following: the idiom of basic objects is the one that we first learn and the one that we normally use to identify and recognize objects not only to others but also to ourselves. So, it is reasonable to suppose that the basic objects have a particular phenomenological salience, in the sense of being introspectively accessible without sustained inspection. Thus, it is also plausible to suppose that the representations that the visual system delivers are representations of the basic objects.

If this is the right way to interpret Fodor, then we have a new proposal concerning what is represented in vision. Like Marr, Fodor thinks of a restricted set of things that are represented by the outputs of visual processing, but for Fodor what is represented are middle-sized objects belonging to basic categories rather than geometrical shapes. Fodor reaches this conclusion by taking people's immediate perceptual reports and their learning behavior as evidence of what is phenomenological salient to them, and, in turn, phenomenological salience as a constraint on the kind of representations we want the modules to output. If he is right, then we have a view that preserves common sense in supposing that we see more than just geometric shapes and in drawing a distinction between what we see and what we can conceive of.

In the next section, I argue that Fodor's argument does not establish the conclusion he wants. Given his own criteria we are bound to accept a richer view of visual content and to admit that how the world appears to us is influenced by the expertise we have.

**The Relativity of Basic Objects**

In the previous section, we looked at Fodor's argument for the view that vision represents more than simple geometric shapes and in particular for the view that vision represents objects belonging to basic categories. His argument is that visual representations should be available to us as premises for conscious decision and inference, so such representations should be easily accessible to us introspectively. Since, the idiom of basic objects is the one that we learn first and the one that we normally use to recognize and identify objects not only to others but also to ourselves it is reasonable to suppose that those objects are phenomenologically salient to us, and this, in turn, suggests that the representations that the visual system delivers are representations of basic objects.

But this argument appears to be open to an obvious objection. What is phenomenologically salient to us, what we are inclined to learn, and what is more natural to report and to recognize in what we observe depends on our aims, interests, competences, and training. Consider phenomenological accessibility first: what is introspectively accessible without sustained inspection in visual perception depends on the aims and expertise we have. If I look out the window and my goal is to see how many trees there are, the trees will be phenomenological salient to me, but this doesn't imply that I see only the trees or that something else couldn't become salient given a change in goal. Furthermore, what is salient to me depends on my previous experience. If something outside the window is particularly familiar, it might catch my attention and become salient. But this doesn't imply that I visually represent only what is familiar or that something else couldn't become salient given a change in my knowledge of my surroundings. Familiarity with things can make them somehow 'stand out' in visual perception: think of seeing a friend in a crowd, her face clearly standing out against the background of other people; or think of spotting a familiar house in a landscape. The house clearly stands out from the background, but one also sees the background.[12]

A similar point applies to learning names of objects or reporting one's visual perceptions. Usually, in learning names and reporting what we see, we want to convey information, and so we tend to learn and use categories that are more useful in doing so. The

[12] In fact, the reverse can also happen: something can become so familiar that one completely ceases to notice it. Think of some of the furniture in your house, that you 'rediscover' only when you have to move. Or think of habituation, the decrease in response to a repeated stimulus: you walk in a room and a sharp odor strikes you immediately. Then after a while, you cease to notice it at all.

level of basic objects is the level of abstraction that allows to convey as much information as possible with as little effort, so it is natural that this is the level that children learn first and the level at which people are inclined to describe what they see. But this does not imply either that we only see the objects that we learn names for and report, or that, if we had other goals and other expertise, we couldn't be more inclined to learn names and to give descriptions using different categories. The fact that people learn certain names and tend to describe what they see in a certain way is due not to the special perceptual status that basic objects enjoy, but to pragmatic considerations. If asked to describe what I see outside my window, for instance, I may do so in terms of trees, cars, and people because this is useful in giving a generic and informative description of what I see. But the fact that I do that, does not imply, first, that I see only what my description conveys: I see cars, trees and people, but I also see their colors, shapes, what's in their background and so on. And secondly, that I couldn't just as promptly give a description in terms of sports cars, pine trees and young people provided that I had the relevant expertise.

That the basic level of categorization can switch depending on one's goals and experience is, in fact, confirmed but some of Rosch's experiments. In one of them a former airplane mechanic is asked to identify various parts of a plane, and his recognitional categories differ significantly from those of common people. Rosch concludes that: "Different amounts of knowledge about objects can change the classification scheme. Thus, experts in some domain of knowledge make use of attributes that are ignored by the average person" (Rosch et al. 1976).

And since Rosch's results were published a large number of psychological and anthropological data have proven that the basic level of categorization can shift given one's expertise and goals. Boster (1986) for example, found that Aguarana women (part of an indigenous population of North Central Peru) who are typically engaged in cultivating manioc[13], tend to refer to manioc plants with highly specific names, while other members of the community who interact less with manioc, name plants at a more general level. This suggests that more specific levels of classification can become basic given one's expertise and aim. Additionally, a more general level of abstraction can become basic as a function of reduced psychological significance: cross-cultural studies comparing US citizens from the Berkeley, California area and the indigenous Maya of Southern Mexico show that for people who depend on plants for their survival, the more specific level (e.g. oak, rather than tree, or even the more specific white oak rather than oak), is most fundamental. Members of cultures for whom plant identification is less vital possess a more inclusive basic level (Atran 1994; Berlin 1972; Berlin 1978; Dougherty 1978).

If this is true, then Fodor's argument for the view that vision represents a restricted set of objects belonging to basic categories is unsatisfactory. By Fodor's own lights, we have to concede that vision represents more, and that what it represents depends on the aims and expertise one has. And this is just what Fodor wanted to deny. What seems to have gotten Fodor into trouble is his initial requirement that the outputs of visual processing be something that is introspectively accessible to the subject and that could serve as a basis for conscious decision and inference. This commitment may seem surprising for an empirically oriented philosopher such as Fodor. Psychologists generally draw a distinction between cognition and conscious thought where cognition is intended to include faculties and mechanisms that work below the level of consciousness. Given this distinction, psychologists

[13] Also, known as Cassava a plant that is extensively cultivated as an annual crop in tropical and subtropical regions for its edible root, a major source of carbohydrates.

tend to suppose that perceptual systems deliver to cognitive systems representations that are not available to conscious thought. In the next section, I consider some possible replies to the line of criticism I developed in this section and argue that they are not successful in rescuing Fodor's view.

**Objections, Replies and Diagnosis**

I argued that Fodor's argument for a restricted set of objects of vision constituted by objects belonging to basic categories is unsatisfactory. The fact that basic objects tend to be phenomenologically salient, that we learn their names early and that we tend to use them to describe what we see does not imply that they are all we see. In fact, given Fodor's own criteria we would have to conclude that what vision represents changes with one's aims and expertise.

Now, Fodor could reply to this kind of criticism by appealing to the notion of modularity, but if my criticism of modularity in chapter two is successful, then we have little reason to suppose that visual systems are modular. Alternatively, Fodor could give up some degree of encapsulation. He may concede, as he sometimes does (Fodor 1988) that long training and expertise could change the classification scheme. Long familiarity and interaction with plants, for instance, could make more specific levels of classification salient. This would amount to giving up some degree of diachronic encapsulation of the modules: but Fodor could claim that complete diachronic encapsulation is not a requirement of his theory. While long training might make a more specific set of objects visually salient, the original set of basic objects might never lose its privileged status. With training, pine trees rather than just trees, might become phenomenologically salient to someone, but trees remain salient too and they are the categories that were first learned. What matters to encapsulation, Fodor might say, is that the modules are not easily affected by short-term and local changes in one's expertise.

But there are at least two problems with this kind of reply: one is that the research on categorization reported above suggests that for people who grow up in radically different cultures more specific levels of classification are, from the beginning, more basic. The category tree is never more salient than the category pine tree or Austrian pine tree. Secondly, there is evidence that identification and recognition are sensitive to short-term and local changes in expertise. People's capacity to accurately identify objects is enhanced when they have been previously exposed to the object. The advantage requires as few as one previous presentation; it lasts at least three weeks, and is found for both words and pictures (Schacter 1987; Goldstone 1998).

If this is true, then Fodor has to concede that visual content is richer than he thinks. What commits Fodor to this conclusion is the idea that the outputs of visual processing should be something that is introspectively accessible to the subject and that could serve as a basis for conscious decision and inference. Given that what is accessible in visual perception changes with one's goals and expertise, Fodor is then bound to conclude that what we see changes with one's goals and expertise, and this is precisely what he wants to deny. What seem to get Fodor into trouble are his epistemological worries. He thinks that the deliverances of visual systems should serve primarily a justificatory purpose: such deliverances have to be available to the subject in order to ground her beliefs and judgments.[14] I come back to this assumption in chapter three.

## Conclusion

The dialectical situation looks pretty grim for the minimalist, and this is something that we started to suspect in chapter two. There, we saw that the minimalist is committed to a set of problematic claims. She wants to hold that vision represents a restricted set of objects and that it represents them as a function merely of contact between our visual system and the environment. At the same time, she wants to hold that visual processes are inferential and so that what they output is partly determined by some previous knowledge of the world. In order to keep what is represented in vision at a respectable minimum, the minimalist has to then commit to modularity. But once we saw that we had little reason to accept modularity, we started to suspect that, as it stands, minimalism cannot be defended no matter how rich it allows visual content to be. And this, it seems, is the conclusion we reached in this chapter.

As a matter of fact, I don't think that the prospects for holding a theory that shares the intuitions that minimalism wants to preserve are so grim. And this is something that we will start to appreciate in the next chapter and then, more fully, in chapter five. By drawing a distinction between visual states that are available to the conscious subject and visual states that are not so available, and by rejecting the view that visual processes are inferential we can preserve the intuition that what vision represents is determined by what kind of visual system we have while also making room for the evidence that shows that how the world appears to

[14] Fodor is explicit about the constitutive link between perception and knowledge also in Fodor et al. (1981).

us is determined by our conceptual resources. But before we move to my positive account we need to consider maximalism. The argument so far indicates that we should be ready to give up the commonsense intuition that there is a clear distinction between perception and cognition. Since we have reasons to suppose that visual systems are not modular and since what is phenomenologically salient in vision can change with a change in expertise, we should be prepared to accept the view that what we see is constitutively determined by our conceptual resources and can change given a change in such resources. And this is just what maximalism holds. The next step, then, is to see what maximalism exactly amounts to, what's involved in adopting it as the correct view of visual content and whether we are bound to accept it. This is the topic of the next chapter.

# IV. MAXIMALISM: WHAT YOU BELIEVE IS WHAT YOU SEE

In chapter two, we saw that visual processing is often described as a kind of inference from low-level sensory representations of the proximal stimulus to high-level representations of the distal causes of the stimulus. The inferences are said to be necessary because the initial data for vision provided by sensory transducers is ambiguous in the sense that a given pattern of light could have been caused by a number of different distal causes. Since, despite this ambiguity, we tend to have a unique percept, the visual system is said to infer the percept by using knowledge of the structure and behavior of objects in the environment. The system organizes the initial information into representations of objects by assuming, for example, that the causes of visual stimulation are rigid and so that the interpretation "rigid structure" should be preferred to a competing interpretation.

In chapters two and three, we saw that accepting this view together with modularity helps to keep what is represented by visual perception at a respectable minimum, and to preserve the idea of a visual base (a set of objects that are perceived independently of our theoretical commitments). Since vision has access to only a limited and proprietary vocabulary in performing the inferences and since this vocabulary is not accessible to, and influenced by, the conscious subject (it is innately specified), the representations that vision outputs are relatively simple and unaffected by one's theory of the world. But since we have reasons for rejecting modularity, we may be inclined to think that visual processes have access to a good deal of stored concepts for performing their operations and that they are affected by the subject's experience. This way of thinking invites two related ideas: one is that what vision represents changes with a change in the stored information one possesses: visual content is plastic.[15] The second idea is that there is no such thing as a visual base, that is, a set of objects that is represented in vision and that is unaffected by the subject's experience. This view is what I called "maximalism": maximalism holds that visual content is constituted by the relations it holds to other contentful items we already possess, paradigmatically concepts. As we will see, maximalists can differ depending on the number of concepts that they allow to be constitutive of visual content.

Maximalism predicts that our immediate perceptual reports and what is phenomenologically salient to us in vision varies with the expertise we have. And this is precisely the conclusion we reached in chapter three. So, we should be inclined to accept that visual content is plastic: doing so is compatible with our intuitions that visual content is relatively rich and that it changes with expertise. But this view leaves no room for the idea of a visual base, a set of objects that is seen independently of one's conception of the world, and it has been used by some philosophers, among them Paul Churchland, to argue for a sort of perceptual relativism that admits too many objects as represented in vision.

In this chapter, I first analyze Churchland's specific brand of maximalism, the reasons and evidence that have been given for it, and the unpalatable consequences that it has. I then show that Churchland's view does not simply follow from accepting that vision involves inferences together with a rejection of modularity. We are then left with a more nuanced kind of maximalism, one that tries to put a limit on the things that can be visually represented while still denying the existence of a visual base. I argue that this brand of maximalism also

[15]The term "plasticity" is actually borrowed from neuroscience where it stands for the idea that neural connections are not fixed.

has unpalatable consequences, and this should motivate us to reject it. I then start suggesting how to do so. Maximalists often appeal to a number of related and mutually supportive facts to argue for their view: one is the evidence that perceptual reports and discriminatory skills vary significantly cross-culturally and depending on one's expertise. I argue that we can avoid taking this evidence as conclusive by drawing a distinction between epistemic and non-epistemic seeing, that is, between visual states that are available to the subject of experience for epistemic purposes and states that are not. I briefly sketch this distinction in this chapter and return to it in chapter six. The second set of facts is constituted by the presence of (what are considered to be) clear examples of plasticity: multi-stable figures, like the duck-rabbit, are clear examples of how visual content can change depending on a change in interpretation. I leave the task of providing an alternative understanding of aspect shifts to chapter five, because doing so involves denying that vision is inferential.

**Radical Plasticity**

If we accept the idea that vision performs inferences but reject modularity we can allow ourselves to think that visual content is fairly rich, and that it is affected by one's expertise. Psychologists who subscribe to this view generally argue that, because of the influence of experience on visual operations, people belonging to radically different cultures and environments might visually perceive the world differently from us (Gregory 1970; Von-Helmholtz 1924-25). People whose surroundings are less populated by geometrical figures, for instance, will be less affected by optical illusions such as the Müller-Lyer because such illusions supposedly rest on interpreting lines as part of three-dimensional geometric objects (Segall et al. 1966).

This kind of view is and has been supported by a plethora of experiments (Bruner et al. 1951; Hastorf 1950; Goldstone 2003; Winawer et al. 2007) aimed at showing that our visual perceptions of color, distance and objects are influenced by what we think, believe, and expect. And this evidence is integrated by neurological evidence that shows back projections of neural activity from areas associated with cognitive faculties (e.g. prefrontal cortex) to the visual cortex (Prinz 2006b). Moreover this view predicts the kind of evidence that we analyzed in the previous chapter. Since visual processing is influenced by one's knowledge of the world, we should expect that what is phenomenologically salient to us and what is most natural to report about what we see will likely depend on the expertise we have. People belonging to different cultures will tend to describe the world differently from us and have visual experiences where different objects are phenomenologically salient.

Paul Churchland, among others, appeals to psychological evidence of this kind to argue for the theory-ladeness and plasticity of perceptual content (Churchland 1979).[16] Sensory states, according to Churchland, are meaningless, and need to be interpreted and exploited by our conceptual resources in order to become meaningful representations of the world.[17] Churchland says:

---

[16]A similar view was first articulated and defended in philosophy by Hanson (1958) and Kuhn (1970).

[17]There is a puzzle, in Churchland, concerning the representations provided by retinal processes that constitute the premises of the non-demonstrative inferences performed in vision. Such representations are supposed to encode information about properties like color and orientation and they are supposed to have their content merely in virtue of the retinal processes they undergo (where such processes are non-inferential). But in *Scientific Realism and the Plasticity of Mind* Churchland insists that a mental state cannot have content unless it is embedded in a theory (where a theory is a network of concepts). This raises doubts as to whether Churchland is entitled to the initial representations given that these are supposed to have content without being embedded in a theory.

> "The meaning of a term (or the identity of a concept) is not determined by the intrinsic quality of whatever sensation happens to prompt its observational use, but by the network of assumptions/beliefs/principles in which it figures. Sensations are just *causal* middle-men in the process of perception, and one kind will serve as well as another so long as it enjoys the right causal connections. (So far then, in principle they might even be dispensed with, so far as the business of learning and theorizing about the world is concerned.)" (Churchland 1979, p.15)

Churchland further claims that virtually any stored representation could be employed in visual processing and so that virtually any concept could figure in the interpretation carried out by the visual system. This is because, according to Churchland, the initial data for vision is incapable of determining which concepts should be mobilized to interpret it. Churchland says:

> "There is nothing written in the nature of things to guarantee, *ab initio* the propriety of the concepts we apply." Churchland (1979, p. 25).

This suggests, first, that what representations we get from visual processing depend on the concepts we already have in the sense that our concepts are constitutive of their content. There is no contentful state we are in that is not conceptually informed. And second, that by acquiring new concepts we can come to interpret our sensory states in new ways and so we can come to perceive the world differently.

What is of particular interest to Churchland is the kind of conceptual resources we acquire when we learn the language of a new scientific theory: learning the Copernican theory, for instance, may produce not only a change in one's beliefs about the shape of the earth, but also a genuine perceptual change in how one visually perceives its surface.

---

Incidentally, Churchland admits elsewhere the existence of a type of content that is the product of mere contact between perception and the world: he calls this type of content "calibrational content" (Churchland et al. 1983). But he seems less willing to accept this type of content in other writings (Churchland 1979 and1989).

Typical examples of perceptual changes that are supposedly due to a change in one's theory are given by perceptual shifts that occur in the presence of ambiguous figures like the famous duck-rabbit extensively discussed by Wittgenstein (Wittgenstein 1953). When one learns a new scientific theory, according to Churchland, one can experience a visual shift of the same kind that one experiences when one goes from seeing the duck-rabbit as a representation of a duck to seeing it as a representation of a rabbit. The content of one's experience and correlatively what is phenomenologically salient to one change given a change in theory. Thus, really well brought up children, according to him, would be able to observe the accretion of atmospheric H2O molecules when they are looking at the gathering of the dew, and listen to aperiodic atmospheric compression waves when listening to the steady roar of the pounding surf (Churchland 1979, p.30).

This way of understanding vision suggests that much of what we can conceive could also be visually perceived, and what is perceived depends constitutively on the kind of expertise one has or has acquired. People belonging to radically different cultures, and in particular people from radically different scientific traditions may see the world differently. Where we see lighting and hear thunder the bronze-age farmer sees "Thor hurling heavenly fire" and hear him "pounding his hammer" (Churchland 1989, p.275). If this is true, then the representations that visual processing output can represent a lot more than just trees, pine trees and plants. They represent things as abstract as molecules and Thor. In fact, although Churchland never explicitly commits to this, it is not clear how the theory could avoid claiming that anything of which we can conceive we can also visually perceive. Given that we can mobilize concepts as far removed from common observation as the concept of a molecule, it is not clear why we couldn't also mobilize the concept of say, a set when looking at groups of objects. The result would be claiming that vision represents not only molecules, but also mathematical entities.[18] If this is true, then there is no clear limit on what we can see: just about anything of which we can conceive we could also visually perceive. In fact, in this picture, there doesn't seem to be any room for a distinction between what we *visually* represent and what we *conceptually* represent.

So, it seems that we have fallen into an opposite pitfall to the one we started off with. We started by looking at a view that seemed too restrictive in allowing only geometrical shapes to be represented in vision. We rejected that view and considered a richer account of visual content. We saw that if we reject modularity and if we take perceptual reports and phenomenological salience as a guide to what vision represents we have reasons to think that visual content is relatively rich and that it changes given a change in expertise. But we now seem to have end up with a view that allows too many things to be represented in vision, and it is not clear whether we have the resources to resist it. Given that the expert chemist could learn to immediately report seeing H2O molecules upon seeing water and that he could perhaps undergo a visual switch where the molecules become salient in her visual experience, we seem to be committed to the view that H2O molecules can be seen. And this would amount to giving up our pre-theoretical intuition that there are things that we can conceive that we cannot perceive. Thus, we need to either find a way to resist this view or learn to live with it.

Learning to live with it involves, at a minimum, the following: people who have significantly different conceptions of the world perceive different things. The world does not appear to them as it appears to us, and the extent to which we perceive the same world is only

[18]Penelope Maddy, by appealing to the same psychological tradition that inspires Churchland, argues that sets can become perceivable if one acquires the relevant set-theoretical expertise (Maddy 1990).

the extent to which we hold similar or overlapping theories. Accepting this claim, threatens the idea that observation serves in the adjudication and resolution of clashes of opinion. Observation is regarded as decisive in resolving disagreements precisely because it constitutes a common ground against which to test theories. This common ground is what radical plasticity about perception threatens to wipe off. What one sees may not be shared by other observers. In the next section, I show that we don't need to be committed to this view.

**Maximalism and Visual Constraints**

Churchland takes the fact that sensory states are meaningless and need to be exploited by our conceptual resources in order to become meaningful representations of the world to suggest that our sensory states could elicit any concept we possess. Any concept, even the concept of molecules, could be used to interpret the sensory data. It can look to us as though H2O molecules are present just as it can look to us as though a tree or a rigid object is present.

Churchland's position may seem to follow from a) the claim that sensory data are ambiguous and b) the claim that visual processes are not modular. But it doesn't: the fact that a given pattern of light on the retina could have been caused by a number (even an infinite number) of different things in the environment does not imply that it can elicit any number of concepts we possess in order to interpret it. Nor does the fact that visual processes are not encapsulated imply that they have access to all of the subject's concepts. We could be "wired up" so as to form always the same representation from a given stimulus (and to form no other representation from that stimulus) even if that stimulus could have been caused by many different things. Indeed, it would be remarkable if natural selection failed to take care of this.

For suppose that a given stimulus can be caused either by a rigid object or by a non-rigid one but, in the environment we live in, it is caused almost always by a rigid object and almost never by a non-rigid one. Then someone who tended to form a representation of a rigid object from that stimulus would presumably have a great advantage over someone who tended to form a representation of a rigid object sometimes and a representation of a non-rigid object other times.

In fact, the reason why psychologists tend to suppose that vision is inferential in the first place is to explain how we form almost always the same percept despite the ambiguity of the initial data. Since, in the vast majority of cases, we see rigid, three-dimensional objects that do not change their shape and size as they move, it is argued that the visual system must know that objects in the environment are like that, and it must use this knowledge in inferring the same percept across encounters with ambiguous data. What psychologists tend to disagree about is whether this knowledge is innate (Fodor 1983, Rock 1983) or acquired from experience (Gregory 1970). If we suppose that it is acquired from experience, then there is a clear sense in which how we see the world varies with such experience: it is because we have previously seen rigid objects that we keep on seeing rigid objects. If we grew up in an environment that did not contain rigid objects we would (perhaps) perceive different kinds of objects.[19]

Churchland, then, seems to overlook the notion of constraint that is often employed in explaining how the visual system limits the interpretations of a given stimulus. We have seen one of these constraints at work in chapter two as part of the explanation of the Muller-Lyer

[19]The goal of cross-cultural studies on the Müller-Lyer illusion is precisely to establish that people living in less "geometric" environments (e.g. the forest) tend to interpret sensory states using different constraints than we do: since the objects they tend to see are less "regular" than the ones we see they should be less affected by the illusion because perceiving the illusion requires seeing the lines as three dimensional angles (Segall et al. 1966).

illusion. Objects in the environment tend to retain their size as they move and as we move around them. But while they retain their size, their retinal projections change in size depending on their distance from the observer. We don't, however, perceive them as changing in size: they look size-constant. The visual system, it is argued, must know that objects in the environment are size-constant and adjust their size depending on their perceived distance. This knowledge, maximalists tend to argue, is not innate (like the modularist supposes) but acquired from experience. It is because we have previously seen that objects are size constant that we keep on seeing them as size constant. Thus, there is a clear sense in which what we see is determined by our previous experience and can change with it.

There is room, then, for holding a more nuanced maximalist position, one that preserves the spirit of maximalism by denying the existence of a visual base but that limits the number of interpretations that can be performed on the sensory data. How to set such limits is likely an empirical issue, one that's left to psychologists to investigate: but since such limits can be set, we can preserve, even as maximalists, the idea that there are things that we can conceive that we cannot visually perceive. Still, I think that even this nuanced version of maximalism is unsatisfactory. In the next section, I try to motivate the importance of preserving a visual base.

**Consequences of Maximalism**

If maximalism is true, then there is no concept- or experience-neutral observation. Any perceptual state we are in is, *qua* state that represents the world in a certain way, also a state that constitutively depends on our concepts. Thus there is no perceptual base, that is, there is no perceptual content that is generated without any influence by concepts.

But there are good reasons for wanting to allow this kind of content. The most significant of these reasons is, I think, the idea that having visual experiences is one important way in which we learn our concepts, so having the experiences should not require already having them. Surely, there are a number of different ways in which we can learn a concept: for instance, there are a number of different ways in which we can learn what a rigid object is. One could learn it from a book, by being told what rigid objects are, by looking at lots of pictures of rigid objects, and so forth. Still, it seems that one important way in which one can learn what rigid objects are is through perceptual encounters with them.[20] And it seems that maximalism cannot make room for this possibility. Since the content of our visual experiences is constitutively determined by our concepts, some concepts are required to have the experiences. In particular, the concept "rigid" is required to have a visual experience of a rigid object.

Maximalists can suppose that the number of concepts that are required to have visual experiences is restricted. They may suppose that we need to posit only a limited number of primitive concepts (paradigmatically colors, rigidity and three-dimensionality). These primitive concepts are what allow us to have visual experiences of a certain kind and such experiences, in turn, ground the acquisition of more complex concepts. But even this nuanced version of maximalism has to concede that some concepts, for instance the concept of

[20]The claim here is neither that having a visual experience of a rigid object is necessary to have a concept of a rigid object, nor that it is sufficient. People may learn what rigid objects are without ever seeing one, and, depending on how robust an account of concepts one holds, a perceptual encounter with a rigid object may not be enough to acquire the concept. The claim is simply that *in some cases* acquiring a concept starts with a perceptual encounter with an instance of the concept, thus having the perceptual encounter should not presuppose already having the concept.

rigidity, are not acquired. Contrary to what we ordinarily think, we don't learn what rigidity is by perceptual encounters with it.

Some maximalists may reply to this point by drawing a distinction between concepts and other mental representations along the lines of Prinz (2002). The idea is that concepts are the kinds of mental representations that can be freely tokened by an observer (e.g. brought to mind) while other kinds of representations may work subpersonally but not be available to the subject of experience. Having a perceptual encounter with a rigid object may only require matching the present percept to a stored representation of a rigid object (Prinz 2006b). But then, one could have a visual experience of a rigid object without yet possessing the concept, by simply matching the present percept with a stored representation of the rigid object previously encountered.

Plainly, this solution does not explain how one acquires the stored representation in the first place. It seems plausible to suppose that we acquire these representations from experience. We start learning what rigid objects are, for example, from perceptual encounters with them: but if maximalism is right, we couldn't have perceptual encounters with rigid objects without already having some representation of them. Thus, maximalism is committed, at a minimum, to the view that we already possess some representations, some very basic knowledge of the world, and that we do not acquire such knowledge from experience.[21]

As I mentioned in the introduction, wanting a theory of visual content that explains how we acquire knowledge of the world from perceptual encounters with it, is not just a

[21]Notice that minimalism is also committed to the existence of some primitive and innate concepts. Since minimalism subscribes to the idea that visual processes are inferential, it is committed to the view that some concepts are a prerequisite to have a visual experience. Specifically, the proprietary concepts that are used to carry out visual inferences: "rigid", "three-dimensional" and so on.

matter of wanting to respect our intuitions. Supposing that seeing the world is one important way in which we learn about it, it is reasonable to want a theory that is at least responsive to the fact of cognitive development. Children and adults differ significantly in the concepts they possess and there is both an order to their acquisition (children master words for objects before

Now, it may be impossible to have such a theory. But, so far as I can tell, we haven't tried all the possibilities. In the next section, I start making room for one such possibility by suggesting a distinction between two kinds of seeing. Drawing this distinction helps to see that variability in reports and in discriminatory skills should not be taken as conclusive evidence for maximalism.

**Perceptual Reports and Discriminatory Skills**

Maximalists are often impressed by the often significant variability in perceptual reports and discriminatory skills both between people belonging to different cultures and between experts and non-experts in a certain domain. Tree experts are faster at discriminating what kind of tree they are looking at than non-experts. They are faster at naming the tree and they are able to categorize trees in more specific (and more general) categories than non-experts. Speakers of Russian who have separate words for blue and light blue, and thus treat light blue as a color rather than as a shade, are faster than English speakers at discriminating light blue from other kinds of blue.

I don't want to underestimate the importance of these results. It seems to me that sincere perceptual reports and discriminatory skills are respectable measures of what is represented in visual perception. In chapter three, we saw that it is precisely because our perceptual reports and what is salient to us in visual experience vary with goals and expertise that Fodor's view is untenable. What we visually represent must be more than just basic objects given that reports and skills can vary so significantly. But in that chapter, I also alluded to a distinction between two kinds of seeing and, correlatively, two kinds of visual content. There is content that is available to the subject of experience and content that is not so available. Drawing this distinction, allows us to see that while the evidence for variability is significant, it is not conclusive for maximalism. The difference between experts and non-experts can just amount to a difference in how available a given representational state is to the subject, rather than to a difference in what the representational state represents.

Surely, our immediate perceptual reports about what we see are affected by our goals and knowledge. If I know nothing about trees, then my immediate perceptual reports will vary significantly from those of a tree expert. And this may be taken to show that there is a sense in which the way the expert sees the world is different from how I see the world. But there is still room for a sense in which we both see the same things. What we report may vary while what we see may be the same. It can, for instance, look to both of us as if the world is composed of a vast array of objects with various shapes and colors. But the expert is able to name the objects, while I am not. In fact, this would seem to explain how I could learn from the expert how to name the objects. We see the same objects and she just tells me how they are called.

Drawing a distinction between what we see and what we are inclined to report (even if immediately and sincerely) makes room for the possibility of holding both that the way the world looks can change, and that it stays the same. And it allows one to be skeptical about the claim that if a scientist were to spontaneously report seeing H2O molecules we would have to ascribe seeing molecules to her.

A similar kind of reasoning can be employed when we consider phenomenological accessibility and discriminatory skills. I take the two to be related since presumably phenomenological accessibility or salience enhances discriminatory abilities. It's because some shades of blue are particularly salient in the experience of Russian speakers that they are better at discriminating the shade. Phenomenological salience can certainly change with a change in expertise. But, again, this need not be taken as a sign that what we see changes with a change in expertise. An object can become salient as a result of expertise: it now looks different. But notice that this is compatible with thinking that the object was represented all along and acquires prominence after the acquisition of expertise.

The moral of this section is that variability in perceptual reports and discriminatory skills, although significant, cannot be counted as conclusive evidence for maximalism. Since we can preserve a sense in which how we see the world is independent of how we describe it or of what objects are salient in it, we can also preserve a sense in which the way we see the world is a function merely of the kind of visual system that we have. There is a distinction between representations that are available to a conscious subject for epistemic purposes and representations that are not so available. Upon looking at a tree one can be visually representing a tree even if one is not aware of such representation and even if one is not able to say that it is a tree one is looking at. In chapter six, I return to this distinction and argue, along the lines of Dretske (1969), for a type of seeing that is non-epistemic in the sense of not being directly connected to cognitive states like beliefs and judgments.

Let me finish this section by noticing that it is best to think of Fodor and Churchland as attempting to give an account of epistemic seeing, that is, an account of the kind of seeing that subjects are aware of, and that is directly connected to doxastic states. This is particularly evident in Fodor's requirement that the outputs of the modules be available to consciousness for justificatory purposes. We saw that thinking of modular outputs this way commits Fodor to the kind of plasticity that he wants to deny. And we saw that by distinguishing between visual states that are connected to other doxastic states and visual states that are not so connected we can make room for an alternative.

Churchland and Fodor exemplify a general tendency, shared by both minimalists and maximalists, to suppose that vision (and perception in general) serves primarily a justificatory purpose. It is in the business of delivering states that we can use to ground our theories. While I don't mean to deny that this is one of the important functions of vision, it seems to me that thinking of it as having a primary function is restrictive. Vision is plausibly concerned with a number of tasks: guiding locomotion, informing memory and so forth. It is not clear that any of these functions is more fundamental than the other, and it is even less clear that in order to perform these different functions vision has to output representations that are introspectively accessible to the subject. If this is a fair criticism, then it is a criticism of Marr's view too. As I have suggested in chapter two, Marr thinks of vision has having the primary purpose of enabling recognition, and this is his main motivation for thinking that visual systems represent object-centered geometric shapes. But if we accept that vision is a multi-purpose tool, then this motivation dissolves.

**Conclusion**

Unlike vision, this chapter had a primary function. It was supposed to make us miss minimalism. Minimalism has certainly its problems but at least it drew a clear line between what we see and what we conceive, and it attempted to preserve independence between the two. If the chapter has failed, I hope that at least it has clarified why one may want to hold on to the idea that how we see the world is partly just determined by the kind of visual system we have. Such idea can make sense of how we learn about the world from our perceptual encounters with it. In the concluding section of the chapter, I made a first step in the direction of preserving this idea. By distinguishing two kinds of seeing and, correlatively, two kinds of content I argued that we can hold on to the view that there is a visual base while also admitting that how we see the world changes with expertise. In the next chapter, I make a second step toward preserving this view. If we deny that visual systems are inferential, we find space for a commonsensical understanding of visual content.

# V. SEEING WITHOUT THINKING

We would like to preserve a set of apparently contrasting intuitions. One is the intuition that our visual world is relatively rich: the world looks to us as a rich panorama of objects instantiating a number of properties. A second intuition is that there is a sense in which how this world appears is a function merely of contact between it and our visual system (provided that the system is working properly and that the conditions are good). It can look to us as though a tree is present even if we don't notice that a tree is present and even if we don't know what a tree is. The third intuition is that there is also a sense in which the way the world appears depends on our knowledge: if we have no conception of what a tree is then it doesn't look to us as though there is a tree.

By now, it should be clear that wanting an account that respects these intuitions is not just a matter of being stubborn about wanting to preserve common sense in the face of scientific discoveries. In the previous three chapters, we looked at theories that deny one or more of these intuitions, and, if my arguments have been convincing, each of these theories is not particularly responsive to the evidence. Minimalism overlooks evidence against modularity and cross-cultural variation in perceptual skills. Maximalism overemphasizes such studies and, in one of its versions, overlooks the notion of constraint employed in vision science. Both are not particularly responsive to evidence concerning our cognitive development.

So, we are back to looking for a theory that respects our seemingly conflicting intuitions. The considerations of the previous chapter, suggest that we can start resolving the apparent conflict by getting clearer on what is meant by "us" in the expression "it looks to us". If we think of the "us" as the conscious epistemic subject, then we can think that the way the world looks to us depends on our knowledge. If, on the contrary, we don't think of the "us" as the conscious epistemic subject we can think that the way the world looks to us is independent of such knowledge. Drawing a distinction between visual representational states that are available to, and understandable by, the subject of experience and visual representational states that are not so available enables us to make room for the idea that although what we report and can discriminate depends on the concepts and expertise we have, what we visually represent doesn't.

But this is only a first step in an account of visual content that respects common sense. We still need to understand how our visual system can produce representational states that are theory-neutral in the sense that their content is determined by mere contact between our visual system and the world. In chapter two, we saw that minimalists try to do this by committing to the idea that visual systems are modular: but given our doubts about modularity this strategy is no longer available. By now, however, the other available strategy should be clear. In chapter two we worried about the compatibility of holding minimalism together with the view that visual systems perform inferences. If we understand visual systems as inferential, then the representations produced by vision are not a function of "mere contact" between our visual system and the environment. They are rather a function of inferential processes that use knowledge of the world to produce representations of it. Thus, the representations, while being the product of visual processing alone, are not free from theoretical commitment. But then it is in the minimalist's best interest to find an alternative explanation of how visual processes take place.

Providing this alternative explanation is the goal of this chapter. In the opening section, I appeal to the notion of "natural constraint" to explain how visual systems can produce representations from ambiguous stimuli without having to ascribe any knowledge of the environment to the system. I then argue that this alternative explanation is as well-suited as the received view to explain cases of misperception, to predict cases of visual agnosia, and to explain visual illusions. I then consider an objection: maximalists often appeal to vision of multi-stable figures, like the duck-rabbit, to argue that visual systems are inferential and affected by the concepts we have. The standard way to understand aspect shifts is to think that the shift is due to a change in the concepts one employs in interpreting the figure: concepts drive the way the figure looks. Since the same stimulus gives rise to different percepts depending on an interpretation, such cases seem to be better predicted by inferential views. I reply to this objection by providing an alternative account of vision in the presence of reversible figures. According to my view, deploying concepts is not necessary to experience a shift. Rather, shifts are due to shifts in attention. By paying attention to different aspects of a figure one can see it differently. I conclude by considering some preliminary consequences of my account for the issue of plasticity.

## Seeing Without Inferences

According to the standard way of understanding visual processes, vision involves an inference from cues to objects. The inferences are brought out by conforming to principles concerning the physical make-up of items in the environment that constrain the interpretations that are allowed. Objects in our environment are rigid, three-dimensional and do not change their shape and size as they move and as we move around them: the visual system assumes that this is the case and acts accordingly by organizing the initial data in a way that is compatible with these assumptions. It then moves from the representation of color; contrast etc. to the representation of rigid, persisting, and three-dimensional objects. This suggests that the visual system already knows, or has some conception, that objects in the environment are usually rigid and three-dimensional and uses such knowledge to interpret the sensory data it receives.

By contrast, think of the visual system as "embodying" (without explicitly representing or drawing inferences from) the principles that it uses to interpret the initial data.[22] The visual system does not need to appeal to the principles: it simply does what it is wired to do, which, as it happens, means working in accordance with the principles discovered by the vision theorist. This means that the system automatically produces a given representation from a given set of inputs by operating in a way that can be described as bringing to bear knowledge about the world, but the system does not itself possess the knowledge. The principles that are said to guide the visual system in performing its transformations are something that *we* ascribe to the system in order to explain how it works: they describe routines according to which the system operates but they are not themselves represented by the system, and so explicitly known by the system. According to this way of looking at vision, the visual system is built in such a way that it produces certain kinds of representations in response to certain kinds of stimuli without having to represent the principles according to which it does so and so without having to possess knowledge about the world. Since the principles are often referred to as "natural constraints", I will say that the

[22] I borrow the term "embodying" from Pylyshyn (1999). My view of how visual processing is carried out is very similar to his, although we disagree on what the outputs of the processing represent.

visual system is "naturally constrained" to act in accordance with certain principles and that it "embodies" rather than explicitly representing those principles.

In order to bring out the contrast between the received view of visual processing and the alternative, think of the different ways in which a system can be said to conform to principles.[23] A system can, first, be said to follow a principle when the system explicitly represents the principle and follows it: like when I explicitly follow my mom's instructions for making Tiramisu'. Next, a system can merely happen to conform to a principle, that is, conforming to a principle may be a mere accident, like when one walks out in pajamas when, unbeknownst to one, it's Halloween. Finally, a system can be said to conform to a principle for a reason, but without representing the principle. Dancing bees are traditionally used as examples of systems that conform to principles for a reason without explicitly representing the principles. Bees dance in their characteristic way for a reason (it is supposedly advantageous to do so) but they don't dance the way they do by following a dancing manual that they internally represent. Bees have supposedly been built to conform to certain principles for dancing. The fact that in describing how they dance we formulate principles that accurately explain the way they dance does not imply that the principles themselves have to be represented by the bees.[24]

Similarly, visual processing can be understood as a rule-governed process where representations of objects are automatically produced in response to the detection of certain

[23]These distinctions can be found in Sellars 1954 and Sellars 1974.

[24]For another example think of the rules of syntax: we are pretty good at judging whether a sentence of our native language is well formed, and we can do so both for sentences we have already encountered and for novel sentences. But it is at least arguable that we are able to do that by explicitly representing and following principles of sentence formation such as "You can get a well-formed sentence of English by taking a Noun Phrase and putting it together with an Auxiliary and a Verb Phrase". The fact that we formulate such principles in order to explain how we make correct judgments about well-formdeness does not imply that we have to attribute an explicit representation of the rules to speakers of a language.

features (color, contrast, etc.) The principles that we ascribe to the visual system in order to explain the kind of representations it generates e.g. the principle that the input should be organized into representations of rigid objects, are not themselves represented by the system. But it is also not an accident that the system works the way it does: it's not by mere chance that we get certain kinds of visual representations. Visual systems are naturally constrained to produce certain outputs given the environment we evolved in: objects in our environment are, usually, three-dimensional, rigid, size-constant, and with stable boundaries, and this explains why we get the kind of visual representations that we do. Because the visual system evolved in our world, the representations it computes are generally (though not necessarily) veridical. For example, because in our world (as opposed to the world of, say, the jellyfish) most objects are rigid, the rigidity constraint will generally lead to veridical perceptions.

Now, there are a number of reasons for thinking that the constraints are built into the system rather than represented by the system. First, their success in producing veridical representations is a function of the environment we evolved in, thus it is preliminarily plausible to think that they were wired in for their evolutionary advantage. Secondly, the constraints are not sensitive to beliefs and judgments. Vision does not respond to any other kind of knowledge or new information related to these constraints (e.g., the constraints show up even if the observer knows that there are conditions in a certain scene that render them invalid in that particular case: remember the Muller-Lyer).

And third, the constraints don't seem to inform other cognitive systems. Observers do not know the principles that enable them to calculate size constancy and rigidity. And even if one takes the view that a natural constraint constitutes "implicit knowledge" not available to consciousness, it is still the case that this knowledge cannot in general be used to draw inferences about the world or that it can be used in any way outside the visual system (Pylyshyn 1999). So, embodying a natural constraint is different from drawing an inference from knowledge of the world (including knowledge of the particular constraint in question).

If this is true, then the visual system (or the subject for that matter) does not need to have any knowledge of rigidity, three-dimensionality and size-constancy in order to produce a representation of a rigid and size-constant object in response to some initial input. Given the detection of some visual features (paradigmatically colors, contrast, orientation etc.) the visual system will automatically produce a representation of a specific rigid object with characteristic shape, color, orientation etc. without having any concept of what rigidity is or what an object is. In this picture, the initial detection of some visual features determines the kind of representations the system produces without having to appeal to an inference. Thus, we can see the world in a certain way by mere contact between our visual system and the environment: in particular, we can see the world as populated by a variety of rigid, persisting objects that have characteristic shapes, colors, and orientations without having any knowledge of what rigid objects are. But then we can have our visual base.

**Received View vs. Natural Constraint View**

In the previous section, I presented an alternative way to understand visual processes that does not make reference to inferences. As I see it, an important motivation for the inferential view of vision is a lack of alternatives and I hope to have provided one. But the alternative has to be shown to be at least as explanatory powerful as the received view. So, in what follows I consider three cases of how the received view and the natural constraint view compare in explaining phenomena that we want a theory of vision to be able to explain.

It is often argued that the received view has an advantage over any other theory of visual processing in explaining how misrepresentation is possible (Fodor and Pylyshyn 1981). Part of an adequate theory of perception ought to be an account of perceptual error, and it is easy to see how error can take place if we adopt the received view. Misperception is connected with failed inferences. Your visual perception that something is a rigid object, for example, is said to depend upon inferences from the "appearance" of the thing (e.g. its surface and contour). The inferences depend upon assumptions about how objects *generally* are. Thus, they are non-demonstrative inferences, and hence fallible. Perceptual error occurs when, for whatever reason, the inferences go wrong.

But a similar strategy is open to the natural constraint view. The visual system is built in such a way that it produces representations of rigid objects whenever doing so is compatible with the proximal stimulus. The system works as it does because it has evolved in our world and objects in our world are *generally* rigid. But since they are not always rigid the system can be "fooled": it can encounter a stimulus that is compatible with the processing of the representation "rigid object" even though the cause of the stimulus is not a rigid object. So, both the received view and the natural constraint view are capable of explaining how misperception is possible.

Moreover, there is puzzle in the received view that is not a puzzle for the natural constraint view: according to the received view, misperception is connected with a failed inference from appearances to objects. Notice that the failure is ascribed to the inferences not to the appearances. In the received view, vision starts with sensory representations that encode information about contrast, acuity, line-orientation and color: such representations are produced automatically by mechanisms of transduction and they are not the product of any inferential process. They are, as it were, the basic appearance of things from which the visual system infers the presence of (rigid and three dimensional) objects. But now we have a dilemma: since the appearances have to serve as premises for inferential processes they have to be representational elements: they have to be elements with conditions of semantic evaluation. But, as representations, they should also be able to misrepresent. Thus the received view has to either a) admit that misperception can be caused not just by failed inference but also by failed "appearance" or b) deny that appearances are representational elements. Given that inferences need premises, accepting b) would amount to giving up the idea that vision is inferential, which is what the natural constraint view wants to do. But if the received view accepts a) then it agrees with the natural constraint view that representations can be produced by non-inferential processes. Either way, the received view has no advantage over the natural constraint view in explaining how misperception is possible.

Next, what I just said suggests that if we accept the natural constraint view we don't need to think of the early stages of vision as representational. Since the primary motivation for thinking of them as representational elements is that they have to serve as premises in inferences, denying that they have to perform such function takes away that motivation. In chapter two, I said that chapter five was going to present further reasons against minimalism1. Now we can appreciate why: if we don't think of vision as inferential we also don't need to think of early visual stages, e.g. stages that encode information about the surface or three-dimensional shape of an object as representational. We don't need to think of vision as first representing "appearances" and then representing objects. Thus, we have one more reason to reject minimalism1. But this leaves open the question of how to explain cases of visual agnosia, not only of associative agnosia, but of any kind of impairment where some complex visual function is impaired while more elementary functions are not (in section 2.4. we briefly saw simultagnosia, a condition characterized by the inability to perceive an object or scene in its entirety while the perception of parts of the object or scene is intact). Inferential accounts of vision may be thought of as having an advantage in explaining these cases since they admit representational levels below that of rigid, three-dimensional objects. But it's not clear why the natural constraint view is not as well suited to understand these cases. Of course, the fact that in cases of *impaired* vision the visual system is able to represent less than what it normally represents does not imply that in cases of *normal* vision it proceeds in stages from less rich representations to richer ones. But additionally the natural constraint view predicts that in impaired vision the visual system will try, unsuccessfully, to produce the representation of an entire object while only being able to represent aspects *of it*.

Finally, consider how the natural constraint view and the received view compare in the explanation of visual illusions. As we saw in chapter two., the Muller-Lyer illusion is often explained by making reference to the assumption of size-constancy. Objects in our environment do not change their size as they move and as we move around them while their retinal projection changes in size depending on their distance from the observer. The visual system is said to assume that objects are size-constant and to adjust the size of the objects depending on how distant they are perceived to be. In the case of the illusion, an inferential analysis would say that the visual system infers the length of the two lines by hypothesizing that they lay at different distances from the observer. By contrast, the natural constraint view says that the visual system is so constructed so that it enlarges the size of items whenever their retinal projection differs in size – i.e. whenever doing so accords with the proximal stimulus. So, the system represents the lines as different in length; this representation, rather than some other possible one, is generated simply because, given the input, and the structure of the system, it is the only one that the system could compute. Notice that this happens independently of knowledge of the particular scene, in fact, despite knowledge to the contrary.

If what I have argued so far is right, then the natural constraint view is as well-suited as the received view to meet a number of desiderata on a theory of vision. But it may seem as though we have been avoiding an obvious objection. The natural constraint view seems well-suited to explain how we always get a *unique* percept from ambiguous stimuli. Since the visual system is built to conform to certain specifiable principles it produces the same output given the same stimulus. But it's then hard to see how this view could make sense of cases of multi-stable perception. Such cases are problematic because the same stimulus produces two (or more) percepts. The received view seems to be in a good position to explain these cases: since a genuine interpretation is involved in vision, the two percepts are produced by a change in interpretation. By switching assumptions, and correlatively, by deploying new concepts, we can see a figure differently. But this kind of explanation is not open to the natural constraint view because the natural constraint view denies that vision involves an interpretation intended as an inference. In the next section, I explain in more detail why cases of multi-stable perception seem to support the idea that vision involves an interpretation. In section four, I that cases of multi-stable perception do not pose a problem for the natural constraint view. I do so by presenting an alternative way of understanding aspect shifts that does not make reference to an interpretation.

**Ambiguous Figures**

Traditionally, cases of vision in the presence of ambiguous figures have been referred to as cases of seeing-as or of seeing-aspects. But their import for a theory of seeing, that is, for a theory of the content of visual perception should be clear. We want to establish what it is that we see in the sense of how the world visually appears to us. But the world always appears to us in one way or another and there is always the possibility of it appearing differently (remember that what is phenomenologically salient in vision can change). A theory of visual content, then, should explain how cases of seeing-as, that is, cases where we switch from one view to another of the same thing, are possible, and how they take place.

Such cases are thought to be better explained by inferential theories of vision and, further, to land support to maximalism. The idea is that switches of aspect should be understood in terms of switching interpretation. Different concepts are used to interpret the figure and this is what gives rise to the two percepts. The concepts the observer has drive the way the figure looks. Maximalists further argue that cases of vision in the presence of reversible figures offer a nice model of what happens in vision in general. An ambiguous stimulus could give rise to multiple percepts: what percept we end up with depends constitutively on the concepts we employ in interpreting it (Hanson 1958).

But the idea that seeing-as involves an interpretation is widely accepted even by non-maximalists (Fodor 2006; Noë 2004). Seeing the duck-rabbit figure as a duck figure, for instance, involves using the concept “duck” to interpret the figure, thus one cannot see the figure as a duck unless one has the concept of a duck and is able to deploy it when looking at the figure. This idea seems to capture particularly well what is involved in undergoing a perceptual shift between two or more ways of seeing the same thing. Since the thing stays the same what changes must be something about the subject and it's natural to think that it is an interpretation. And the view seems well equipped to explain a number of intuitive features of seeing-as. Reversing an ambiguous figure is often under the observer's control and it "feels like" moving from one interpretation to another of a figure or object. [25] Think of looking at clouds and being able to see them in multiple ways. In so doing, we seem to be able to switch from one view to another of the clouds by actively giving a new interpretation, and so by subsuming what is seen under new concepts.

Finally, evidence from psychology lands support to this way of understanding seeing-as. In line with the tradition that invokes 'central dynamic factors' such as imagining and interpreting in order to explain reversals (Von Helmholtz 1924-25) Irvin Rock hypothesizes that knowledge of the ambiguity and knowledge of the terms of the ambiguity are necessary to experience a reversal. One is able to experience a shift from seeing a duck figure to seeing a rabbit figure only if one knows that the figure is ambiguous, and only if one knows what the terms of the ambiguity are. If this view of reversals is correct, then it confirms that one needs to know what concepts to deploy in order to experience a perceptual shift because concept deployment is what is responsible for them.

In a number of studies Rock attempts to confirm this hypothesis (Girgus et al. 1977; Rock et al. 1992; Rock et al. 1994; Gopnik et al. 2001). The studies generally involve subjects that are unacquainted with ambiguous figures, and they test observation under two

[25] Churchland (1989), Kuhn (1970) and Hanson (1958) all stress the voluntary control that subjects have over perceptual shifts suggesting, at times, that interpreting is not only necessary but also sufficient to experience a shift. The latter claim has come under considerable attack, for instance by Fodor (1988), and the evidence does not support it. So, in this section, I will concentrate primarily on the issue of whether an interpretation is *necessary* to experience a shift.

conditions. In the first condition, subjects are uninformed of the ambiguity and of its terms; they are simply asked to look at a figure and to report what they see.[26] In the informed condition, subjects are asked again to look at the figure after being informed of its ambiguity and of its terms.[27] Rock found that, in adults, roughly 60% of observers never reverse in the uninformed condition and that all of them (100% in one study and 99% in another) do in the informed condition. Further, in order to avoid previous familiarity with ambiguous figures (which might account for 40% of spontaneous reversals in adults) Rock tested children between the age of 3 and 4, and found that no child reverses in the uninformed condition, while 44% reverses in the informed condition.

These results suggest that knowing that a figure is ambiguous and knowing the terms of the ambiguity is a good predictor of whether a subject will experience a reversal. And this, in turn, indicates that reversals involve an action of interpretation on the part of the subject. So, it seems that we have good reasons to accept the idea that seeing-as involves an interpretation. This view accounts for a number of intuitive features of seeing-as and it is supported by evidence on bi-stable perception. In the next section I show that there is an alternative way to understand seeing-as does not make reference to an interpretation and that is therefore compatible with non-inferential accounts of vision.

## Seeing-As without Concepts

[26]The instructions are: "Tell me what you see. Continue to look at the picture since I am going to ask you some questions about it later." (Girgus et al. 1977 p. 41).

[27]The informed condition is preceded by an interview. During the interview, subjects are told that the figures they are seeing can be seen in multiple ways and they are informed of the terms of the ambiguity: this is done by showing them unambiguous versions of the ambiguous figures they have seen.

According to my alternative understanding, seeing one thing as another only involves seeing something in a way that is driven by what visible features one pays attention to. Seeing a figure as a duck-figure, for instance, involves seeing the figure *in a certain way*, and an interpretation is not required to see something in a certain way. We see the figure as a duck figure because the figure is duck-shaped, and not because we think of it that way.

The first step in arguing for this alternative view is to question Rock's experiments presented in the previous section. The experiments seem to raise a pressing question: we need to understand why there is a large gap between children and adults in experiencing reversals. Why is it that, in the informed condition, almost all the adults can reverse while more than half the children cannot? Proponents of the view that seeing-as involves an interpretation may explain this fact in, roughly, two ways: either children do not yet have a concept of what ducks and rabbits are, or they have such concepts but they are not good at applying them.

Suppose, first, that we appealed to a fairly robust notion of what concept possession involves: children aged four, even if possessing language and even if able to recognize and identify ducks and rabbits upon encounter do not yet possess concepts of them. Children this age may not have the knowledge required to have independent thoughts about ducks and rabbits. Thus, the evidence that shows that young children are unable to reverse ambiguous figures even when informed may be explained by the fact that young children lack concepts.

This way of preserving the view that seeing-as involves an interpretation, however, requires some independent motivation on pain of just being *ad hoc*. Since young children exhibit the kind of abilities that are plausibly associated with concept ascription, we need an *independent* reason for withholding such ascription, that is, a reason that is not constituted *merely* by the desire to preserve the view.

Moreover, this way of arguing is in tension with evidence on bi-stable perception in non-human animals. If four year-olds cannot be credited with concepts, then it seems that non-human animals, that lack language, cannot be credited with concepts either. Thus, non-human animals should, according to this way of arguing, be unable to experience reversals. Now, although there aren't many studies on bi-stable perception in non-human animals,[28] the studies that we do have suggest that animals like pigeons are capable of reversals.

But now suppose that we adopt the second alternative. We deflate our requirements on concept ascription and we willingly ascribe concepts to pigeons and young children. The failure of four year olds to reverse ambiguous figures when informed is due not to their lack of concepts but to their lack in the ability to apply them. Young children are not yet able to interpret a figure. A fairly popular psychological theory supports this view. According to the theory (Gopnik et al. 2001; Sobel et al. 2005; Mitroff et al. 2006) young children are unable to reverse because they lack a theory of mind, that is, they lack an understanding that the mind is an interpretative device, and so that the same visual stimulus can be interpreted in different ways by different observers (Gopnik et al. 1988). In a number of studies, Gopnik attempts to show that the ability to reverse ambiguous figures is correlated with the emergence of the ability to pass false beliefs tasks and to grasp the distinction between

[28] As far as I know there are only a few experiments involving bi-stable perception in non-humans. There are a small number of studies on binocular rivalry in cats and monkeys, and there is a study on reversible figures in pigeons (Sengpiel et al. 1995; Sheinberg et al. 1997; Vetter et al. 2000). The study on pigeons tested their ability to experience perceptual shifts (Vetter et al. 2000). Eight experimentally naïve pigeons were trained to discriminate the apparent horizontal motion of two dots appearing on a display from their apparent vertical motion. Pigeons indicated which motion they perceived by pecking on a key. After the training period, pigeons were shown an ambiguous display where vertical and horizontal motion alternate: their pecking behavior suggests that they all experienced reversals between the perception of the two motions.

appearance and reality. Children younger than five seem unable to understand that others may have different beliefs than their own, and that one's own belief may be mistaken. Gopnik argues that children are able to experience reversals only after they start grasping the idea that multiple interpretations of the same figure are possible.

The accuracy of this theory is still very much under consideration, but a few studies on autistic children cast a shadow on the possibility of its success. Autistic children are famously bad at theory of mind tasks, having difficulty in understanding the distinction between appearance and reality, and the difference between their beliefs and those of other people. Yet, when tested on ambiguous figures they seem to have no more problems than normal children in reversing them, especially when informed of the ambiguity (Sobel et al. 2005).

This brings us back to the question of understanding why there is such a difference between children and adults in reversing. It seems to me that the empirical evidence does suggest why children at four cannot reverse ambiguous figures. The ability to reverse seems to emerge fairly consistently at age five (Mitroff et al. 2006)[29] and it is highly correlated with the capacity to direct and hold attention to different features of a figure or object. By focusing on different features, different kinds of visual operations are prompted that give rise to different percepts. The reason why young children cannot experience a reversal is because the development of attentional mechanisms, and of the capacity to control them, is a relatively late achievement: pre-frontal cortex which is generally regarded as the *locus* of such mechanisms, develops relatively late in humans reaching full maturation only during adolescence (Diamond, 2002).

[29] When presented with the duck/rabbit or Rubin's vase/face, 35% of children aged 5 through 9 can reverse even when uninformed; and by age 10, 47% spontaneously reverses (Sobel et al. 2005).

A number of studies lend support to the idea that attentional mechanisms are essentially involved in interpreting ambiguous figures. Neuro-imaging studies confirm the stimulation of pre-frontal areas during reversals (Kleinschmidt et al. 1998) and patients with unilateral frontal brain damage have greater difficulty in shifting from one aspect to the other of a reversible figure than normal control subjects (Ricci et al. 1990). Research on bilingual children further supports this view: bilingual children at 6 years of age are more likely to experience reversals than their monolingual peers (Bialystok et al. 2005).[30] This result can be attributed to the fact that bilingual children develop control over selective attention earlier than monolingual children because they have to control two active language systems (Bialystok 2001 and Bialystok et al. 2004).

In adults, the part of a reversible figure that a subject focuses on has been shown to determine which percept the viewer experiences (Chastain et al. 1975). In one experiment, subjects were presented with segments of the ambiguous rat-man: some segments were expected to produce the perception of a rat while others the perception of a man. The experiment confirmed that the starting segment determines the perception of the figure, and so that paying attention to some features rather than others determines which percept the viewer experiences.[31]

Given this evidence, we have reasons to believe that the formation of a given percept of an ambiguous figure results from focusing attention on a focal area that contains features significant for one percept but not for the alternative one. Focusing on the mouth of the duck-

---

[30]In the experiment both bilingual and monolingual children are partly informed of the ambiguity, that is, they are told that a given figure (duck/rabbit, rat/man, vase/face) can be seen in two different ways but they are not told (until the very end of the session) what the terms of the ambiguity are (Bialystok et al. 2005).

[31]This result was confirmed in a second study: subjects focusing on the area that is predicted to facilitate one of the percepts before the presentation of a figure, tended to report the corresponding perceptual interpretation (Tsal et al. 1985).

rabbit figure, for instance, gives rise to the perception of the rabbit shape, while focusing on the beak favors the perception of a duck shape. Thus, reversing ambiguous figures requires being able to focus and hold attention on some relevant parts of a figure.

If this is true, then it suggests a preliminary idea for understanding seeing-as in a way that does not explicitly appeal to an interpretation or to concept application: seeing-as is seeing something in a way that is determined by the features one pays attention to. Reflection on other ambiguous figures suggests that this preliminary view is plausible. Think of the Necker cube: it seems that in the case of the Necker cube, one switches from one to the other view of the cube in a pretty mechanical way by focusing attention on different angles of the cube and without having to deploy concepts such as 'three-dimensional" and "cube".

The defender of the standard view may reply that directing attention requires concepts. You need to know what to direct your attention to in order to do it (remember Plato's famous adage: "How will you look for it Socrates, when you do not know at all what it is? How will you aim to search for something you do not know at all?") But vision scientists focusing on attention generally agree that attentional mechanisms are driven partly by intrinsic biases of the perceptual system toward certain kinds of stimuli (Remington 1980; Desimone et al. 1995; Blaser et al. 1999). Local in homogeneities (like a single red item in a field of grey items), new objects, and objects that are larger, brighter and faster moving are among the things that capture one's attention automatically, and quite independently of the subject's conceptual development (Desimone et al. 1995). What one learns how to do, is to have some control over attentional mechanisms, and so to select information that is relevant for the performance of a given task.

Furthermore, even when attention is under the subject's control, directing it to the relevant features of a figure or object may not require a concept of what one is looking for. For one thing, one may just *happen* to be paying attention to a relevant feature: remember the data from Rock. In the uninformed condition adults can reverse 44% of the time and children at five can reverse 35% of the time. Why couldn't this happen just as a result of a fortuitous focus on the relevant features of a figure without appeal to concept application?

In the present dialectical situation appealing to the inferential nature of vision is not open to the defender of the view that seeing-as involves an interpretation. For, whether vision is inferential is precisely what is at issue. And if it is true that directing attention does not require concepts, then we can think of aspect-shifts in the following way: by specifically paying attention to certain features of an object or figure the visual system detects certain colors, contrasts etc. It then processes the input into representations that, given the principles it has been built to follow, are determined by what features have been detected initially. By stumbling upon the features of a figure that favor a given percept, the visual system automatically delivers a representation of a specific item even though the subject may lack a concept of what the item is. By paying attention to one angle of the Necker cube, for instance, the visual system forms a representation of a three-dimensional object positioned in a certain way even though the system and the subject may lack a concept of three-dimensionality.

Similarly, by paying attention to the beak-shaped portion of the duck-rabbit figure the visual system produces a representation of a specific item with a characteristic shape, color and orientation. The item has a characteristic "look". What kind of look? The look of ducks. Ducks, like much else, look a certain way quite independently of whether we perceive them to look that way (Austin 1964; Jackson 1977). The figure can simply look like a duck: it can be duck-shaped. Thus, by paying attention to the relevant feature of the figure one can see something duck-shaped without yet possessing the concept of a duck (or of shape). Similarly, by paying attention to the features of the figure that favor the perception of a rabbit the visual system automatically produces a representation of something that looks different: it has, for one thing, a different orientation. Thus, something else can be seen in the figure, and, as it turns out, it resembles a rabbit: it looks like one (at least the face of one).

If this is a sensible way to understand aspect shifts, then we have reasons to suppose that a subject can see something *in a certain way* without possessing a concept of it. So, a subject can also see a figure as a duck figure without possessing a concept of a figure or of a duck. And this view seems well-fitted to explain the intuitive features of seeing-as that the standard view also explains: seeing-as is often under the subject's control because subjects generally have control over what features of a figure or object they pay attention to. Further, the alternative is compatible with the view that once we have concepts, we can use them as a guide to what we focus on. Thus, when told that a figure can look like a duck we can search for the beak-shaped part of the figure and prompt visual processes that deliver a representation of something duck-shaped. Seeing-as, then, feels like subsuming something under a new description simply because giving a description guides what we pay attention to, and this, in turn, is what prompts visual operations that give rise to different percepts.

Moreover, according to the alternative, seeing one thing as another involves a kind of visual curiosity where attention is paid to different features of a figure or object.[32] This

[32]Richard Wollheim (Wollheim 1980) talks of seeing-as as a kind of visual curiosity and distinguishes it from seeing-in that involves more imaginative abilities. I agree with him that there is an interesting distinction to be drawn here but I have no space to expand on it. Let me just notice that this distinction is what motivates my care in saying that when we see the duck-rabbit figure switch, we see it switch from a duck *figure* to a rabbit *figure*.

requires that a subject is able to engage in visual search where attention is focused on different features. If one is not able to engage in visual search, then her chances of reversing will be greatly reduced. This explains, I think, why children younger than five who are unable to focus and hold attention cannot reverse. Additionally, if seeing-as calls for engaging in visual search then subjects who are unaware that visual search is what they are supposed to be doing may reverse much less frequently. When uninformed, subjects may fail to reverse because they are just staring at the same point in a figure, or they are moving their attention aimlessly from one point to another. So, the alternative seems well-equipped to explain the evidence provided by Rock.

If what I have argued so far is right, then my alternative way of understanding seeing-as is at least as plausible as the dominant view according to which seeing-as essentially involves an interpretation. But if this is true, then cases of multi-stable perception cannot be used to argue in favor of inferential accounts of vision. A non-inferential account explains them just as well. Moreover, cases of multi-stable perception cannot be used to argue in favor of maximalism: since an interpretation is not required to experience reversals such cases show us precisely how to get visual content that is free from theoretical commitment. They show us how something can look to us in a certain way even if we lack concepts. In the next section, I explain some preliminary consequences for the plasticity of visual content.

## The Natural Constraint View and Plasticity

---

We don't, strictly speaking, see the figure as a duck or as a rabbit. For, a duck and a rabbit are not there to be seen and seeing-as is a matter of finding out, by paying attention, what is there to be seen.

If the account of visual perception I defended in this chapter is plausible then seeing is highly constrained. In the vast majority of cases, given a certain stimulus, the visual system automatically produces a single representation. And we can expect that similar stimuli will produce similar representations. This accounts for the relative constancy of our visual world across encounters. The tree I see today outside my window looks very similar to the tree I saw yesterday outside my window. And this is also true of cases of ambiguous figures. Since the stimulus is the same, the representations produced by paying attention to different parts of the figure resemble one another. The duck figure and the rabbit figure look alike: they are somewhat "roundish" and they are both outlines. The two percepts produced by the Necker cube resemble each other in both shape and color. They have some visible features in common. In general, the different representations that an ambiguous figure generates resemble each other in shape or color or orientation.

If this is true, we are in a better position to see why the scientist, just like us, is not able to see the world in terms of molecules and light waves no matter how sophisticated her knowledge of them is. Aspect shifts, like visual perception in general, are constrained by what features are detected initially and by the principles that govern visual operations. If visual operations start with the detection of certain shapes and colors they will end representing objects with those specific shapes and colors. But then there will be no way to switch from seeing an item with a specific shape and color, say a glass of water, to seeing molecules; for the two don't look alike. Even supposing that molecules have a specific look they don't look like water.[33]

[33]A similar kind of reasoning can be employed in the case of Thor. Even if Thor had a characteristic look (which I suppose we would derive from pictorial representations of Thor) he would not look like lightning: he would not have a similar shape or color. So, looking at lightning we would not be able to have a perceptual shift that resulted in a visual representation of Thor.

But we can, of course, experience a shift from, say, seeing a tree to seeing a pine tree; for trees and pine trees look alike. In chapter three, we characterized these cases as cases where an object becomes phenomenologically salient in one's experience. A scene containing pine trees may look different before and after we know what pine trees are. The preliminary suggestion to understand these cases is that paying attention to a previously seen tree can make it salient in one's experience. Vision scientists often say that attention can cause an object to acquire "distinctiveness" partly because what we pay attention to is also generally where we foveate. Paying attention to the features of a tree that make it distinctively a pine tree rather than some other kind of tree (the shape of its trunk, the needle-like foliage etc.) can make it "stand out" in visual perception. The way that expertise influences this process is by directing attention to the relevant features of the tree. The difference between us and experts in a certain domain, then, is not a difference in what we represent: there is a sense in which the world looks to us as it looks to them just in virtue of sharing similar visual systems. It's rather that experts know what to look for: they know what features to pay attention to. And, as I suggest in the next chapter, this is what explains their discriminatory advantage.

**Conclusion**

Theories that subscribe to an inferential view of vision (both minimalists and maximalists) explain how it is possible for something to look rigid or three-dimensional to us, by appeal to some previous knowledge (either innate or acquired) of what rigidity and three-

The situation for mathematical entities seems even worse. For, supposedly, mathematical entities like sets and numbers don't look like anything. Since they have no looks, they can't be detected by our visual apparatus.

dimensionality are. The natural constraint view and the account of seeing-as that I defended in this chapter reverts this order of explanation. Things look rigid and three-dimensional to us *not* because we (or our visual system) know that they are, but because they *are* rigid and three-dimensional. Since they are rigid and three-dimensional we evolved to represent them as such.

So, it seems that we have found a visual base. The world appears to us as populated by a vast array of objects with characteristic shapes, colors, orientations, and positions: and it appears to us that way merely in virtue of contact between the world and our visual system. But then we are in a good position to accommodate two of our pre-theoretical intuitions: the idea that we see a relatively wide range of objects, and the idea that we see them merely in virtue of the visual system that we have. We are left with the task of making room for the third intuition: the idea that the way the world appears partly depends on our knowledge. In the last section of this chapter, we looked at a sense in which this is true. By acquiring knowledge one can learn to pay attention to the relevant features of a figure or object, that is, to those features that make it distinctively the object that it is. Being able to do so favors the development of discriminatory skills which are often associated with perceptual plasticity. In the next chapter, I expand my analysis of perceptual plasticity and offer a model of what's involved in learning to visually represent the world in new ways. By doing so, I hope to complete the task of rescuing common sense.

# VI. Perceptual Development

The aim of the previous chapter was to rescue two of our commonsensical intuitions. One is the intuition that our visual world is relatively rich, comprising more than, say, just shapes and colors. The other is that how the world appears to us is partly just a function of the visual system that we have. The aim of this chapter is to explain how we can make room for the third intuition: the idea that how we see the world can change given a change in knowledge and expertise. If this can be done, then it will also serve to further support the view that our visual world is particularly rich comprising meaningful objects and their properties.

Traditional accounts of visual content, both minimalists and maximalists tend to explain perceptual development in terms of conceptual development. We learn to see the world differently when, by acquiring new concepts we can deploy them in interpreting our visual states. Concepts work like categories: by learning the concept "pine tree" we can subsume a visual representation of a tree under the new category. By contrast, the account I start developing in this chapter maintains that perceptual development occurs when a given representational state becomes progressively available to the subject of experience for epistemic purposes. This availability is a matter of the integration of the representational state in a network of others. Since integration is a matter of degree, perceptual development is also a matter of degree. It involves the progressive development of certain abilities. I will talk specifically of three (there are more, but these seem to capture important stages of development): one is the ability to discriminate, tell apart or distinguish (I use these terms interchangeably). The other is the ability to identify or recognize (I also use these two terms interchangeably). And the third is the ability to name or report.

Making room for the idea that we learn to see the world differently requires first drawing a distinction, along the lines of Dretske (1969) between epistemic and non-epistemic seeing. I draw this distinction in the opening section of this chapter and suggest that perceptual development can be understood as moving from one kind of seeing to the other. By borrowing from developmental psychology, I then suggest a model of what's involved in moving from one perceptual state to the other, and explain that this move is more complex than a simple passage from representations in analog form to representations in digital form (the analog/digital distinction is embraced by Dretske). I then reply to an objection: I am trying to show that the range of things we (epistemically) see can change with a change in expertise. Minimalists may claim that epistemically seeing is not really a kind of seeing. There can be, they argue, some development in how the world appears given what we see, but the development is fully conceptual. Thus, we can quickly represent a pine tree given that we see a tree but this representation is a conceptual one: strictly speaking, we do not visually represent pine trees. I provide some reasons for thinking that epistemic seeing is a genuine kind of seeing and so that learning how to see epistemically is genuine perceptual development.

**Epistemic and Non-Epistemic Seeing**

In the previous chapter, I argued that perceptual processes, rather than being inferential processes are naturally constrained processes that automatically produce representations of quite specific three-dimensional objects with specific shapes and colors (specific looks) even though the subject may lack the resources to appreciate such objects. The observer may lack the concept of an object or may lack the concept of what specific object she is looking at. In fact, the subject may even be unaware that she perceives an object at all. Visual systems make contact with cognitive faculties at multiple levels. Their outputs may be stored in memory without making it to conscious thought.

This way of looking at visual representations is reminiscent of the kind of phenomenon Fred Dretske calls 'non-epistemic seeing' (Dretske 1969). There is a type of seeing, Dretske argues, that we have in common with a fly. This type of seeing does not entail that the subject has any particular belief or set of beliefs about that which is seen, or in Dretske's terminology, this type of seeing has "no positive belief content". I can see a tree without believing that I am seeing a tree, or believing that I am seeing a plant, or believing that I am seeing anything at all. No particular belief is essential to the seeing.
The examples that Dretske proposes in support of his view are similar to some of the examples I used in arguing for the need to keep cognition and conscious thought separate. Dretske says:

> I have occasionally been in such a preoccupied state that as I walked down the street I was, in the only way I can think to describe it, unaware of anything around me. It was only after I snapped out of the 'fog' that I realized I hade been seeing certain things without being aware of it; that is, I can remember having seen things, but I cannot remember being aware, at the time I was seeing them that I was seeing something or that things were looking a certain way to me." (11)

In a second example Dretske describes a case where an ascription of a visual state is justified merely by the fact that a subject has a well functioning visual system and is staring at something in good light conditions:

> The physical and physiological conditions were such that the object must have looked some way to him. 'You must have seen that cuff link; you were staring right at it.' Whatever response this allegation may prompt, it is not refuted by an appeal to ignorance; 'I did not notice it', or 'The drawer looked empty to me.' He may have seen the cuff link without noticing it...What would refute the allegation is something quite different. 'No, I did not look into the drawer', or 'No, I had my eyes closed all the time'...We can still insist that he must have seen the cuff link, and we can attempt to establish our point by leading him back to the drawer for a second look. This may or may not decide the issue. If he exclaims, upon looking into the drawer for the second time 'Oh, here it is; I guess I wasn't thinking the last time I looked, ' it is likely, although the case is by no means clear-cut, that he did see it on the previous occasion, but as we sometimes say, it did not *register*...We must resort to indirect methods arguing by analogy that since, in all likelihood, the conditions affecting the visibility of the cuff link are the same now as when we formerly looked into the drawer, and since it is clear that he now sees the cuff link, it is plausible to suppose that he also saw it beforehand without, of course, realizing it. (21).

In line with these examples, the view of visual processing I defended in the previous chapter holds that one can have a visual perception of something of which one is unaware and for which one lacks a concept. When looking at a tree, for instance, one is (if the conditions of viewing are acceptable) in a visual representational state that represents a specific three-dimensional object with a specific look. This specific three-dimensional object is, say, a tree, and a pine tree and a mountain pine tree. So, there is a sense of seeing, the non-epistemic sense, in which one can see a tree, a pine tree and a mountain pine tree without having any knowledge of what trees, pine trees and mountain pine trees are, and without even being aware of the tree. In other words, there is a sense of seeing in which the content of our visual representations depends entirely on the kinds of objects that are in the world and on the contact they make with our visual system.

This way of understanding visual perception has the advantage of explaining cases like the following: suppose that you know nothing about trees but are asked to cut down all the pine trees in a yard (this example is modified from Siegel 2006). You have to learn some relevant facts about pine trees in order to perform your job, and once you do that, you learn to discriminate and recognize pine trees very quickly. Now, suppose that after you learn to discriminate and recognize pine trees you remember having seen them before. You remember having seen pine trees when you knew nothing about them. This, it seems to me, is evidence that you visually represented pine trees in the past even if you couldn't have discriminated them or recognized them. You visually perceived pine trees because you have a certain visual system that, given the detection of certain shapes, colors, orientations, and positions, produces representations of specific three dimensional objects that can happen to be pine trees. This, you can do even without possessing concepts and even without being aware of what you are seeing.

Assuming that this is true, it seems also plausible to think that there is a sense in which one does not see a pine tree if one cannot discriminate it. This is the epistemic sense of seeing: the epistemic sense of seeing has it that one does not see something unless seeing it makes a difference to her conscious epistemic life.[34] There is a sense in which it does not look to you as though a pine tree is present unless you have some discriminatory abilities when it comes to pine trees (and the 'you' here is the conscious subject): if you can't tell a pine tree apart from other kinds of trees then you don't see one.

Epistemic seeing is a fairly complex phenomenon. Like Dretske's examples suggest it requires, at a minimum, that the subject is aware of seeing something (Dretske usually speaks of "noticing"). But as I understand it here, being aware is not sufficient for seeing epistemically. Seeing epistemically plausibly involves also discriminating, recognizing, and, at a further level of complexity, reporting. These are all different skills but they seem to mark

[34] A visual representation can also make a difference to someone's unconscious epistemic life. It can prompt the formation of perceptual beliefs that we just find ourselves with. Being in such a representational state would not count as epistemic seeing, since the subject is not aware of it.

different stages in what are generally recognized to be perceptual abilities. They also usually ground our ascriptions of visual states to ourselves and others. Discriminating, as I am using it here, is an ability that precedes recognizing and naming: it amounts to simply being able to distinguish one thing from another by sight. One may be able to tell a pine tree apart from an oak tree upon looking, but not be able to recognize a pine tree on further occasions. Discriminating is plausibly the minimum requirement to ascribe epistemic seeing to ourselves and others. This is brought out particularly well by thinking of shades of color: if we cannot tell two shades of color apart, then there is a sense in which we do not see them. It does not look to us as though two shades are present.

Recognizing is plausibly more complex than discriminating: in fact, it seems to presuppose discriminating. On traditional accounts of visual perception, recognizing is a matter of matching the present percept with a stored representation (section 2.3.) As I will argue shortly, we should rather think of recognizing as connecting a representation to one's knowledge base, with no need for matching. Whether or not we are able to recognize something is also generally taken as a measure of whether we saw it epistemically. If you did not recognize a pine tree upon encounter, then it did not look to you as though a pine tree is present. Finally, I am thinking of both discriminating and recognizing as independent of language. Young children and creatures that do not posses language can discriminate and recognize. Thus, I take it young children and animals are capable of some degree of seeing epistemically. Still, being able to report what one sees may be considered a sign of epistemic seeing (although it is not required for it). Being able, for instance, to immediately report seeing a mountain pine tree, on top of a tree and of a simple pine tree, is a sign that perceptual development has occurred, that is, that the tree looks different to the observer.

The suggestion, then, is that the epistemic sense of seeing is also the sense in which one can learn to visually perceive new things by acquiring expertise. In section 5.4., we saw that theoretical knowledge can enhance one's discriminatory abilities by directing attention to the relevant features of an object. Doing so, makes the object salient in one's experience and salience is usually a prerequisite for discrimination. The acquisition of new discriminatory abilities may in turn prompt the acquisition of recognitional abilities and such abilities make a difference to our conscious epistemic life.

Dretske regards the difference between epistemic and non-epistemic seeing as a difference between a perceptual and a cognitive state. This difference, according to him, is a coding difference between states that are encoded in analog form and states that are encoded in digital form. Perception and cognition make contact, according to Dretske, when the information in analog form delivered by perceptual systems is converted in digital form for selective use by cognitive faculties. By contrast, I want to suggest that the difference between epistemic and non-epistemic seeing is a difference between two kinds of perceptual states: the fact that epistemic seeing can be influenced (in a way that needs clarifying) by our expertise, does not affect its status as a perceptual state.[35]

There are a number of reasons why I think that epistemically seeing a pine tree is being essentially in a perceptual state rather than in a cognitive state. I present these reasons later in this chapter. What I shall do next is suggest a model for understanding what is involved in moving from non-epistemically seeing something to epistemically seeing it. I

[35]My account may be compatible with Dretske's criterion, but Dretske seems to imply a change in representation akin to conceptualization, and I do not support this. I aim at supplementing the approach in Dretske by drawing a further distinction, lost on his account, between concepts and high-level perceptual representations that are not yet conceptual.

suggest that doing so involves a number of phases where a representation becomes more and more accessible to the subject and integrated in a network of others.

**From Non-Epistemic to Epistemic Seeing**

Perceptual development is usually motivated by task-demands. As in the example of the previous section, we may need to learn to discriminate, recognize and name pine trees for practical purposes. If what's been said so far is right, we already have some material to start with. The world appears to us as containing objects with characteristic looks even if we don't know what the objects are. Since, epistemic seeing requires, at a minimum, that one is aware of an object development is often initiated by paying attention to some objects rather than others that are relevant to the performance of a task. By paying attention, an organism can bring certain objects to prominence in visual perception, and this favors the development of certain discriminatory abilities. Cognitive scientists generally speak of objects to which we attend as "acquiring distinctiveness" or "being emphasized" (Honey et al. 1989). Discriminatory abilities can then be acquired by simply observing an object.

But discriminatory abilities can also be enhanced by learning facts about objects, in particular facts that can direct our attention to the relevant features of it. Discrimination is generally eased if we know what to look for. Importantly, learning facts about an object can also, so as to say, situate an object in a network of others. It can connect our idea of a pine tree, to other ideas: pine trees have needle-like foliage, have a soft, moist, inner bark, can be used as survival food, and are evergreen and resinous. Acquiring knowledge of a certain object can aide perceptual development in two ways: it can direct our attention and it can integrate a representation in a network of others that we (presumably) already have. It can tell us what a particular three-dimensional object we are looking at is and how it is related to other objects. This integration facilitates recognition. Recognizing an object involves being in a visual state that is connected to our knowledge base in the sense that the object represented can be immediately placed in relation to others we already know about. Correlatively, failure of recognition happens when we cannot place an object that's visually represented in relation to others. Now, the more accessible and integrated a given representation is, the more liable it is to verbal report when language is present.

The suggestion, then, borrowed from developmental psychology, is that perceptual development consists in a visual representation becoming available and integrated. The representation becomes data *to* the subject rather than just being data *in* the subject. (Karmiloff-Smith 1992).[36] One learns to visually represent the world in new ways when one's visual representation is available for conscious discrimination, recognition and reporting. To give a sense of what it means for a representation to be integrated, we can appeal to an example from cognitive development (Clark et al. 1993). When children between 4 and 11 years old are asked to produce drawings of usual and unusual objects their abilities vary significantly. By roughly 41/2 years old children can rapidly draw a man. Their drawing procedure is fairly fluent: they first draw the head, then the body and then the limbs. This contrasts with their ability to draw less basic objects, for instance a man that does not exist. Younger children, although they announce that they are going to draw a pretend man, proceed to use their usual sequence to produce a normal man. They cannot manipulate the

[36] Karmiloff-Smith does not talk specifically about perceptual development. She is primarily interested in conceptual and linguistic development. I borrow from her the idea that development consists in the progressive accessibility and integration of a given representation. Representations that are initially unavailable for use by the subject become progressively available.

component parts of the representation of a man they already possess. Striking developmental differences emerge, however, between the 4-6 years old group and the 8-10 years old group. The modifications made by the younger children involve changes in values of variable of size and shape and certain types of deletion. Older children's modifications are different in nature. They insert extra elements from same conceptual category (extra legs) or from different categories (extra wings) and they can change the orientation or position of the parts. This evidence is suggestive that older children have more integrated and thus more cognitively flexible representations than younger ones. Similarly a visual representation can be more or less integrated in a network of others upon being produced by the visual system.

The idea, then, is that seeing something epistemically involves being in a visual representational state that is accessible to the subject and integrated in a network of others. If this is true, then learning to see the world is something that may happen in stages because accessibility also happens in stages: it involves acquiring some discriminatory, recognitional and linguistic skills. This means that the extent to which someone sees something epistemically is a matter of degree: one may be able to discriminate something upon looking without yet being able to recognize it, and one may be able to discriminate and recognize something without being able to name it. I take this fact to explain why there can be disagreement about whether to ascribe a visual representational state to someone (and to ourselves). In the epistemic sense, whether or not it looks to you as if a pine tree is present depends on whether you can tell a pine tree apart from another tree, recognize it, and report seeing it, but since these skills don't need to come in a package, we can disagree about whether or not you are in that visual representational state. This view has the advantage to explain why we may have doubts about ascribing visual representational states to developing children who may exhibit some, but not all, of the skills that we associate with epistemic seeing.

If what I have said so far is plausible, then the difference between non-epistemic and epistemic seeing is primarily a difference in the kinds of skills that are associated with each. It is not a difference in what we see but a difference in what we can do given what we see. Non-epistemically seeing something does not require that we are able to discriminate, recognize or report that which we see. By contrast, epistemically seeing something involves at least being aware and discriminating that which we see and can also include recognizing and naming it. This account allows us to explain the intuitions that often motivate maximalism. Maximalists tend to suppose that what discriminatory, recognitional and linguistic skills one exhibits when looking at something are fundamental in deciding whether to ascribe a given visual representational state to them. If one cannot discriminate, recognize or report what one sees then one shouldn't count as visually representing it. The opposite is also generally true: if one can discriminate, recognize and report something one can be credited with visually representing it. Maximalists think that this is the case because what we visually represent depends constitutively on the concepts and expertise we have and since concepts and expertise come with certain abilities, display of such abilities is a necessary condition on the ascription of visual content. While this is not exactly true in the view I favor, there is a sense in which we can preserve these intuitions: whether or not someone displays certain abilities is fundamental in deciding whether they epistemically see something, even though it is not fundamental in deciding whether they see something at all. So, drawing a distinction between epistemic and non-epistemic seeing, allows us to preserve the idea that people possessing similar visual systems can share visual contents while also accommodating maximalists' intuitions. What changes from person to person is the availability and integration of a given content and so the kinds of discriminatory and recognitional skills that come with such availability. This, I take it, is the sense in which people can inhabit different visual worlds while also sharing the same visual world.

## Is the Development Perceptual?

The account of perceptual development I sketched in the previous section may raise the question of whether it is an account of *perceptual* as opposed to conceptual development. I have been trying to show that there is a sense in which we can come to perceive new things by paying attention and acquiring expertise in a certain domain. What this development involves is the development of discriminatory, recognitional and linguistic abilities. But one might object that the kind of development I outlined is a type of *conceptual* development done on what one already sees: one can come to know what pine trees, oak trees etc. are, that is, one can come to possess a concept of them but one doesn't, strictly speaking, see them. In this section, I want to provide some reasons for thinking that the development is perceptual and so that epistemic seeing, although influenced by one's expertise in the way described above is a kind of *seeing* rather than a kind of thinking. Doing so involves giving some platitudes on how to think of perceptual rather than conceptual states.

The first reason why I think of seeing epistemically as a perceptual state is given directly by the idea that we can see epistemically (just as we can see non-epistemically) only objects that have characteristic looks. Our visual system is able to represent items that have characteristic shapes, colors, orientations, and position. I have left the notion of what a characteristic look is, fairly vague. Expanding on it would exceed the scope of the present work. And there are certainly questions concerning what counts as having a characteristic look: objects vary significantly in shape, color, orientation and position (think of trees). Whether or not something has a characteristic look depends on whether or not it has a look that distinguishes it from others: for instance, trees have a characteristic look if they have a general shape, color, orientation and position that distinguish them from chairs, cars, people etc. But so far as objects have a characteristic look, it is plausible to think that they are visually represented rather than inferred from, or thought about, given what is visually represented. So, if we think that pine trees, and oak trees have characteristic looks, then it is also plausible to think of them as perceivable.

Secondly, the skills associated with epistemic seeing are usually taken to be paradigms of perceptual skills. Being able to quickly discriminate and easily recognize an object is generally considered a mark of perceptual rather than conceptual abilities. This is what makes the data on variations in discriminatory skills and perceptual reports both across cultures and across levels of expertise compelling. Recall Rosch's experiments: if the idiom of middle-sized objects belonging to basic categories (cars, trees, chairs etc.) is the one that we commonly employ in identifying objects not only to others but also to ourselves, then it is plausible to suppose that those objects are perceived by subjects rather than inferred from what is perceived. And since Rosch's experiments also measured reaction times in identification tasks (the time it takes to recognize an object as, say, a tree) the experiments seem particularly apt to showing that middle-sized objects belonging to basic categories are epistemically perceived. Now, since a change in training and expertise (a training that can be as brief as an experimental session) can change one's recognitional abilities, what's epistemically seen can also change. Airplane mechanics epistemically see parts of planes that are not epistemically seen by a non-expert: they are not accessible to non-experts. And in cultures that rely on the identification of specific plants for their survival, people epistemically see plants that we do not epistemically see. We are blind to aspects of the environment that they epistemically (Boster 1986, Atran 1994). Notice that none of this commits us to maximalism. It is perfectly ok to non-epistemically see something while also being epistemically blind to it.

If this is convincing, then the view I favor enlarges the number of things that can be seen. There are two senses in which our visual world is rich. In the non-epistemic sense, our visual world comprises a vast variety of objects that have characteristic looks. *Contra* minimalism1, we don't just see shapes. In the epistemic sense, our visual world also comprises a variety of objects with characteristic looks that we can consciously discriminate, recognize and name: in this sense, the visual world can contract and expand with the contraction and expansion of our skills. *Contra* Minimalism2 our (epistemic) visual world is not fixed. Visual content is, in a sense, plastic: what one is able to discriminate, recognize, and report changes.

## Conclusion

Traditional accounts of visual content, both minimalists and maximalists tend to explain perceptual development in terms of conceptual development. We learn to see the world differently when, by acquiring new concepts we can deploy them in interpreting our visual states. By contrast, the account I started developing in this chapter maintains that perceptual development occurs when a given representational state becomes progressively available to the subject of experience for epistemic purposes. I have suggested that when this happens we move from non-epistemically seeing something to epistemically seeing it. Epistemically seeing something amounts to visually representing it, being aware of it, and being able to discriminate, recognize or name what is seen. Non-epistemically seeing something requires only visually representing it. If this is a sensible distinction, then we are in position to see in what sense the way the world appears to us changes with our knowledge and expertise. Since our discriminatory, recognitional and linguistic skills change with knowledge and expertise what we (epistemically) see also changes. And this makes room for our last intuition that our conception of the world and what we think of it can change how we see it.

# VII. Conclusion

We started out by considering a set of apparently contrasting intuitions that we ascribed to common sense. One is the intuition that our visual world is relatively rich: the world looks to us as a rich panorama of objects instantiating a number of properties. A second intuition is that there is a sense in which how this world appears is a function merely of contact between it and our visual system (provided that the system is working properly and that the conditions are good). It can look to us as though a tree is present even if we don't notice that a tree is present and even if we don't know what a tree is. The third intuition is that there is also a sense in which the way the world appears to us depends on our knowledge: if we have no conception of what a tree is, then it doesn't look to us as though there a tree is present.

We gave some reasons for thinking that wanting a theory of visual content that respects these intuitions is not just a matter of wanting to preserve common sense, but also a matter of wanting a theory that is responsive to the evidence in cognitive science. We then started looking for such a theory but failed to find one. We analyzed two broad types of contemporary theories that essentially disagree on whether there is a visual base, that is, a sense in which how the world appears to us is a function merely of contact between it and our properly functioning visual systems. Minimalism admits a perceptual base but is then unable to account for how changes in concepts and expertise cause perceptual changes. Maximalism denies the existence of a perceptual base but has difficulty explaining how people with significantly different conceptions of the world can see the same things. We also saw that the arguments for each view are unsatisfactory. Since both views share substantial assumptions we then questioned the assumptions.

One is the assumption that visual processes are inferential processes where knowledge of the world is brought to bear to produce a percept from ambiguous visual data. We argued that visual processes can be understood, instead, as naturally constrained processes that automatically produce percepts from a given stimulus. Understanding visual processes this way allow us to make room for a visual base, a set of objects that are seen independently of one’s theoretical commitment. The other is the assumption that visual systems have a primary function, in particular the function of providing percepts that are available to the organism for justificatory purposes. We argued against this assumption and showed that there is a sensible distinction to be drawn between percepts that are available to observers for epistemic purposes and percepts that are not so available. By drawing this distinction we made room for the idea that how the world appears to us can change with our knowledge. Further, denying both of these assumptions made sense of why our visual world is relatively rich comprising a variety of objects that instantiate a number of properties. If the arguments have been successful, we now have a theory of visual content that preserves common sense while also being responsive to the evidence in cognitive science.

# BIBLIOGRAPHY

## Books and Articles

Atran, S. (1994) "Core Domains Versus Scientific Theories: Evidence from Systematics and Itza-Maya folk biology," in L.A. Hirschfeld & S.A. Gelman (Eds.) *Maping the Mind: Domain Specificity in Cognition and Culture*, New York, Cambridge University Press.

Austin, J.L. (1964) *Sense and Sensibilia*, New York, Oxford University Press.

Berlin, B. (1972) "Speculations on the Growth of Ethno Botanical Nomenclature," *Journal of Language and Society*, 1, pp. 51-86.

Berlin, B. (1978) "Ethno Biological Classification," in Rosch, E. & Lloyd, B.B. (Eds.) *Cognition and Categorization*, Hillsdale, NJ: Erlbaum, pp. 9-26.

Bialystok, E. (2001) *Bilingualism in Development: Language, Literacy, and Cognition*, Cambridge, UK, Cambridge University Press.

Bialystok, E. & Martin, M. (2004) "Attention and Inhibition in Bilingual Children: Evidence from the Dimensional Change Card Sort Task," *Developmental Science*, 7, pp. 325-339.

Bialystok, E. & Shapero D. (2005) "Ambiguous Benefits: the Effect of Bilingualism on Reversing Ambiguous Figures," *Developmental Science*, Nov; 8(6), pp. 595-604.

Blaser, E. Sperling, G. & Lu, Z.L. (1999) "Measuring the Amplification of Attention" *Proceedings of the National Academy of Science, USA*, vol. 96, pp. 11681-11686, September.

Boster, J.S. (1986) "Exchange of Varieties and Information between Aguarana Manioc Cultivators," *American Anthropologist*, 88, pp. 428-436.

Brandom, R. (1994) *Making it Explicit: Reasoning, Representing and Discursive Commitment*, Cambridge, Mass.: Harvard University Press.

Bruner, J. Postman, L. & Rodrigues, J. (1951) "Expectation and the Perception of Color" *American Journal of Psychology*, Apr. 64 (2) 216–227.

Bruner, J.S. (1957) "On Perceptual Readiness," *Psychological Review, 64,* 123-152.

Büchel, C. & Friston, K.J. (1998) "Dynamic Changes in Effective Connectivity Characterized by Variable Parameter Regression and Kalman Filtering," *Human Brain Mapping*, 6, pp. 403-408.

Chastain, G. & Burnham, C.A. (1975) “The First Glimpse Determines the Perception of an Ambiguous Figure”, *Perception and Psychophysics*, 17 (3), 221-224.

Churchland, P. (1979) *Scientific Realism and the Plasticity of Mind*, Cambridge University Press.

Churchland, P. (1989) “Perceptual Plasticity and Theoretical Neutrality: a Reply to Jerry Fodor,” in P. Churchland, *A Neurocomputational Perspective: the Nature of Mind and the Structure of Science*, A Bradford Book, the MIT Press, Cambridge, Massachusetts.

Churchland, P.S. & Churchland, P.M. (1983) “Stalking the Wild Epistemic Engine,” *No&ucirc:s*, Vol. 17, No.1, A.P.A. Western Division Meetings, pp. 5-18.

Clark, A. & Karmiloff-Smith, A. (1993) “The Cognizer’s Innards: A Psychological and Philosophical Perspective on the Development of Thought,” *Mind & Language*, Vol. 8 No. 4.

Davidson, D. (1975) “Thought and Talk” in D. Davidson, *Inquiries into Truth and Interpretation*, Clarendon Press, Oxford, Oxford University Press, 2001.

Desimone, R. & Duncan, J. (1995) “Neural Mechanisms of Selective Visual Attention,” *Annual Review of Neuroscience*, 18, pp. 193-222.

Diamond, A. (2002) “Normal Development of Pre-Frontal Cortex from Birth to Young Adulthood: Cognitive Functions, Anatomy, and Biochemistry,” in D. Stuss & R. Knight (Eds.) *Principles of Frontal Lobe Functioning* (pp. 466-503). New York: Oxford University Press.

Dretske, F. (1969) *Seeing and Knowing*, Chicago: University of Chicago Press.

Dougherty, J.W.D. (1978) “Salience and Relativity in Classification,” *American Ethnologist*, 3, 1978, pp. 66-80.

Ellis, A.W. & Young, A.W. (1988) *Human Cognitive Neuropsychology*, Hillsdale, NJ: Erlbaum.

Fahle, M. & Morgan, M. (1996) “No Transfer of Perceptual Learning Between Similar Stimuli in the Same Retinal Position,” *Current Biology*, 6, pp. 292-297.

Farah, M.J., Peronnet, F., Gonon, M.A. & Giard, M.H. (1988) “Electrophysiological Evidence for a Shared Representational Medium for Visual Images and Visual Percepts,” *Journal of Experimental Pychology. General,* Vol. 117, No. 3, pp. 248-257.

Farah, M.J. (2004) *Visual Agnosia*, Second Edition, A Bradford Book, the MIT Press, Cambridge Massachusetts and London, England.

Fodor, J.A. & Pylyshyn, Z.W. (1981) "How Direct is Visual Perception? Some Reflections on Gibson's 'Ecological Approach'" *Cognition*, 9, pp. 139-196.

Fodor, J.A. (1983) *Modularity of Mind*, A Bradford Book, the MIT Press, Cambridge, Massachusetts, and London, England.

Fodor, J.A. (1988) "A Reply to Churchland's "Perceptual Plasticity and Theoretical Neutrality"," *Philosophy of Science*, Vol. 55, No. 2 (Jun, 1988), pp. 188-198.

Fodor, J.A. (1998) *Concepts: Where Cognitive Science Went Wrong*, Oxford: Clarendon Press, New York: Oxford University Press.

Fodor, J.A. (2003) *Hume Variations*, Oxford: Clarendon Press; New York: Oxford University Press.

Fodor, J.A. (2006) "The Revenge of the Given" unpublished. Available on-line at http://www.nyu.edu/gsas/dept/philo/courses/representation/papers/Fodor.pdf

Frisby, J.P. (1980) *Seeing: Illusion, Brain and Mind* New York: Oxford University Press.

Gauthier, I. Behrmann, M. & Tarr, M. (1999) "Can Face Recognition really be Dissociated from Object Recognition?" *Journal of Cognitive Neuroscience*, 11, pp. 349-371.

Gauthier, I., Skularski, P., Gore, J.C. & Anderson, A.W. (2000) 'Expertise for Cars and Birds Recruits Brain Areas Involved in Face Recognition," *Nature Neuroscience*, 3, pp. 191-197.

Gibson, J.J. (1979) *The Ecological Approach to Visual Perception*, Boston: Houghton Mifflin.

Girgus, J. Rock, I. & Egatz, R. (1977) "The Effect of Knowledge of Reversibility on the Reversibility of Ambiguous Figures," *Perception and Psychophysics*, Vol. 22(6), 550-556.

Goldstone, R.L. (1998) "Perceptual Learning," *Annual Review of Psychology*. 49: pp. 585-612.

Goldstone, R.L. (2003) "Learning to Perceive while Perceiving to Learn," in R. Kimchi, M. Behrmann, & C. Olson (Eds.) *Perceptual Organization in Vision: Behavioral and Neural Perspectives,* New Jersey: Lawrence Erlbaum Associates, pp. 233-278.

Gopnik, A. & Astington, J. W. (1988) "Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction," *Child Development*, 59, pp. 26-37.

Gopnik, A. & Rosati, A. (2001) "Duck or Rabbit? Reversing Ambiguous Figures and Understanding Ambiguous Representations," *Developmental Science*, 4 (2), 175-183.

Gregory, R.L. (1970) *The Intelligent Eye*, New York, McGraw-Hill.

Hanson, N.R. (1958) *Patterns of Discovery; an Inquiry Into the Conceptual Foundations of Science*, Cambridge University Press.

Hastorf, A. (1950) "The Influence of Suggestion on the Relationship Between Stimulus Size and Perceived Distance" *Journal of Psychology, 29,* 195-217.

Helmholtz, H. (1924-1925) *Treatise on Physiological Optics*, translated from the 3d German ed. Edited by James P.C. Southall, Rochester, N.Y.] The Optical Society of America.

Honey, R.C., & Hall, G. (1989) "Acquired Equivalence and Distinctiveness of Cues," *Journal of Experimental Psychology*, 34: pp. 87-103.

Ihde , D. (1977) *Experimental Phenomenology: an Introduction*, New York, Putnam.

Jackson, F. (1977) *Perception: a Representative Theory*, Cambridge [ENG.]; New York: Cambridge University Press.

Jiang, Y., Haxby, J.V., Martin, A., Ungerleider, L.G., Parasuraman, R. (2000) "Complementary Neural Mechanisms for Tracking Items in Human Working Memory," *Science*, 287, pp. 643-646.

Johnson, K.E., & Mervis, C.B. (1997) "Effects of Varying Levels of Expertise on the Basic Level of Categorization," *Journal of Experimental Psychology: General*, 126, pp. 248-277.

Kanwisher, N. (2000) "Domain specificity in Face Perception" *Nature Neuroscience*, vol. 3 no 8.

Karmiloff-Smith, A. (1992) *Beyond Modularity: A Developmental Perspective on Cognitive Science*, A Bradford Book, The MIT Press.

Kleinschmidt, A., Buchel, C., Zeki, S. & Frackowiak, R.S.J. (1998) "Human Brain Activity During Spontaneously Reversing Perception of Ambiguous Figures." *Proc. R. Soc. Lond. B. Biol. Sci.* 265, pp. 2427–2433.

Kuhn, T. S. (1970) *The Structure of Scientific Revolutions*, Chicago, University of Chicago Press.

Maddy, P. (1990) *Realism in Mathematics*, Oxford, Clarendon Press.

Marr, D. (1982) *Vision: a Computational Investigation into the Human Representation and Processing of Visual Information*, W.H. Freeman and Company, New York.

McDowell, J. (1996) *Mind and World*, Harvard University Press.

Mitroff, S. R., Sobel, D.M. & Gopnik, A. (2006) "Reversing how to Think about Ambiguous Figure Reversals: Spontaneous Alternating by Uninformed Observers," *Perception*, vol. 35, pages 709-715.

Noë, A. (2004) *Action in Perception*, Cambridge, Mass.: MIT Press.

Peacocke, C. (1992) *A Study of Concepts*, Cambridge, Mass.: MIT Press.

Price, H.H. (1950) *Perception*, London: Methuen.

Prinz, J.J. (2002) *Furnishing the Mind: Concepts and their Perceptual Basis*, Bradford Books: the MIT Press, Cambridge, MA.

Prinz, J.J. (2006a). "It the Mind Really Modular?" in R. J. Stainton (Ed.) *Contemporary Debates in Cognitive Science* Oxford: Blackwell.

Prinz, J.J. (2006b) "The Content of Sensation and Perception," in T. Gendler and J. Hawthorne (Eds.), *Perceptual Experience*, Claredon Press, Oxford.

Purves, D. & Andrews, T.J. (1997) "The Perception of Transparent Three-Dimensional Objects," *Proceedings of the National Academy of Science, USA*, Vol. 94, pp. 6517-6522, June, Neurobiology.

Pylyshyn, Z. (1999) "Is Vision Continuous with Cognition? The Case for Cognitive Impenetrability of Visual Perception," *Behavioral and Brain Sciences*, Vol. 22, No. 3, Jan. pp. 341- 423.

Remington, R.W. (1980) "Attention and Saccadic Eye-movements" *Journal of Experimental Psychology: Human Perception and Performance*, 6, pp. 726-744.

Ricci, C. & Blundo, C. (1990) "Perception of Ambiguous Figures after Focal Brain Lesion" *Neuropsychologia*, 28, pp. 1163-1173.

Rock, I. (1975) *An Introduction to Perception*, New York, Macmillan.

Rock, I. (1983) *The Logic of Perception*, A Bradford Books, MIT Press, Cambridge MA.

Rock, I. & Mitchener, K. (1992) “Further Evidence of Failure of Reversal of Ambiguous Figures by Uninformed Subjects,” *Perception*, 21, 39-45.

Rock, I. Gopnik, A. & Hall, S. (1994) “Do Young Children Reverse Ambiguous Figures?” *Perception*, 23, 635-644.

Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M. & Boyes-Braem, P. (1976) “Basic Objects in Natural Categories,” *Cognitive Psychology*, 8, 382-439.

Rosch, E. (1978) “Principles of Categorization” in E. Rosch & B.B. Lloyd, *Cognition and Categorization*, Hillsdale, New Jersey: Lawrence Erlbaum.

Russell, B. (1912) *The Problems of Philosophy*, London; New York, Oxford University Press.

Schacter, D.L. (1987) “Implicit Memory: History, and Current Status,” *Journal of Experimental Psychology: Learning, Memory, and Cognition* 13, 501-518.

Segall, M.H., Campbell, D.T., & Harskovits, M. J. (1966) *The Influence of Culture on Visual Perception: An Advanced Study in Psychology and Anthropology*, The Bobbs-Merrill Company Inc., Indianapolis, New York.

Sellars, W. (1954) “Some Reflections on Language Games,” *Philosophy of Science*, 21: 3, pp. 204-228.

Selllars, W. (1956) “Empiricism and the Philosophy of Mind,” in *The Foundations of Science and the Concepts of Psychoanalysis, Minnesota Studies in the Philosophy of Science,* Vol. I, ed. By H. Feigl & M. Scriven, University of Minnesota Press; Minneapolis, MN, pp. 127-96.

Sellars, W. (1974) “Meaning as Functional Classification,” *Synthese* 27, pp. 417-37.

Sengpiel, F., Blakemore, C., & Harrad, R. (1995) “Interocular Suppression in the Primary Visual Cortex: a Possible Neural Basis of Binocular Rivalry,” *Vision Research*, 35, 179-195.

Sheinberg, D.L. & Logothetis, N.K. (1997) “The Role of Temporal Cortical Areas in Perceptual Organization,” *Proceedings of the National Academy of Science USA*, Vol. 94, pp. 3408-3413, April.

Siegel, S. (2006) “What Properties are Represented in Perception?” in T. Gendler and J. Hawthorne (Eds.), *Perceptual Experience*, Claredon Press, Oxford..

Sobel, D. M., Capps, L. M. & Gopnik, A. (2005) “Ambiguous Figure Perception and Theory of Mind Understanding in Children with Autistic Spectrum Disorders,” *British Journal of Developmental Psychology*, 23, pp. 159-174.

Spelke, E.S. (1990) “Principles of Object Perception,” *Cognitive Science*, 14, pp.29-56.

Tardif, T. (1996) “Nouns are not Always Learned Before Verbs: Evidence from Mandarin Speakers’ Early Vocabularies,” *Developmental Psychology*, 32: pp. 492-504.

Thornton, I.M. & Fernandez-Duque, D. (2000) “An Implicit Measure of Undetected Change,” *Spatial Vision*, 14, pp. 21-24.

Tomasello, M. Kruger, A.C. & Ratner, H.H. (1993) “Cultural learning” *Behavioural and Brain Sciences,* 16: pp. 495-552.

Tsal, Y. & Kolbert, L. (1985) “Disambiguating Ambiguous Figures by Selective Attention” *The Quarterly Journal of Experimental Psychology*, 37, (A), pp. 25-37.

Turnbull, O.H. Driver, J. & McCarthy, R.A. (2004) “2D but no 3D:Pictorial-Depth Deficits in a Case of Visual Agnosia,” *Cortex*, 40, pp. 723-738.

Ullman, S. (1979) *The Interpretation of Visual Motion*, Cambridge, Mass.: MIT Press.

Uttal, W.R. (2001) *The New Phrenology: the Limits of Localizing Cognitive Processes in the Brain*, Cambridge, Mass.: MIT Press.

Vetter, G. Haynes, J.D. & Pfaff, S. (2000) “Evidence for Multistability in the Visual Perception of Pigeons” *Vision Research*, 40, 2177-2186.

Winawer, J., Witthoft, N., Frank, M., Wu,, L., Wade, A. & Borodutsky, L. (2007) “Russian Blues Reveal Effects of Language on Color Discrimination,” *PNAS*, vol.104, no.19, pp. 7780-7785.

Wittgenstein, L. (1953) *Philosophical Investigations* translated by G.E.M. Anscombe, Blackwell Publishers.

Wollheim, R. (1980) *Art and its Objects: with Six Supplementary Essays* Cambridge, New York; Cambridge University Press.